J. Nonlinear Sci. Vol. 9: pp. 255–332 (1999)

Journal of Nonlinear Science

Delay Embeddings for Forced Systems. I. Deterministic Forcing

J. Stark

Centre for Nonlinear Dynamics and its Applications, University College London, Gower Street, London, WC1E 6BT, United Kingdom

Received March 13, 1995; final revision received April 3, 1998; accepted April 21, 1998
Communicated by Stephen Wiggins

Summary. Takens Embedding Theorem forms the basis of virtually all approaches to the analysis of time series generated by nonlinear deterministic dynamical systems. It typically allows us to reconstruct an unknown dynamical system that gives rise to a given observed scalar time series simply by constructing a new state space out of successive values of the time series. This provides the theoretical foundation for many popular techniques, including those for the measurement of fractal dimensions and Liapunov exponents, for the prediction of future behaviour, for noise reduction and signal separation, and most recently for control and targeting. Current versions of Takens Theorem assume that the underlying system is autonomous. Unfortunately this is not the case for many real systems; in the laboratory we often force an experimental system in order for it to exhibit interesting behaviour, whilst in the case of naturally occurring systems it is very rare for us to be able to isolate the system to ensure that there are no external influences. In this paper we therefore prove two versions of Takens Theorem relevant to forced systems: one applicable to the case where the forcing is unknown, and the other to the situation where we are able to determine independently the state of the forcing system (usually because we are responsible for the forcing ourselves). In a subsequent paper we shall show how to extend these results to give an analogue of Takens Theorem for randomly forced systems, leading to a new framework for the analysis of time series arising from nonlinear stochastic systems.

1. Introduction

Takens Embedding Theorem provides the theoretical foundation for the analysis of time series generated by nonlinear deterministic dynamical systems. Since its publication in 1980, it has stimulated a vast range of applications in fields ranging from fluid dynamics through electrical engineering, to biology, medicine, and economics. In particular, it has led to both a re-examination of old data sets and the construction of new experiments, with the aim of detecting and perhaps even taking advantage of deterministic behaviour in time series that were previously thought to be random (for a good overview, see e.g. Ott et al. [1994]). One might thus even say that this one theorem has given rise to virtually a new branch of nonlinear dynamics, often informally called *chaotic time series analysis*.

Informally, Takens Theorem says that if we take a scalar observable φ of the state x of a deterministic dynamical system, then *typically* we can reconstruct a copy of the original system by considering blocks $(\varphi(x_t), \varphi(x_{t+\tau}), \varphi(x_{t+2\tau}), \ldots, \varphi(x_{t+(d-1)\tau}))$ of d successive observations of φ, for d sufficiently large. Here x_t is the state of the system at time t, and $\tau > 0$ is some sampling interval.

For Takens Theorem to be valid, we need to assume that both the dynamics and the observations are autonomous, that is, independent of time and of any outside influence. In particular, we suppose that there is some mapping f such that $x_{t+\tau} = f(x_t)$, and that φ depends on x only.

Unfortunately, this is not the case in many real systems. Thus, in the laboratory we often force an experimental system in order to elicit interesting behaviour, whilst in the case of naturally occurring systems, it is very rare for us to be able to isolate the system to ensure that there are no external influences. Thus many real systems are often best modelled by equations of the form

$$\begin{aligned} x_{t+\tau} &= f(x_t, y_t), \\ y_{t+\tau} &= g(y_t), \end{aligned}$$

where x_t, as before, represents the state of the system we are interested in, and y_t describes the state of some forcing system. The best known example of this is the case of periodic forcing with some period $T > 0$. In this case y_t is given by the phase $\theta_t = t/T \pmod 1$ of the forcing at time t and g is simply a rigid rotation of the unit circle

$$\begin{aligned} x_{t+\tau} &= f(x_t, \theta_t), \\ \theta_{t+\tau} &= \theta_t + \tau/T \qquad \pmod 1. \end{aligned}$$

At first sight, it might seem that we can apply Takens Theorem to forced systems by enlarging our concept of the state of the system to (x_t, y_t). The dynamics is then given by the pair (f, g), and our observation of the system is given by the function $\psi(x_t, y_t) = \varphi(x_t)$.

The problem with this argument is the word *typical* in the statement of Takens Theorem. In particular, the theorem does not state that given any (f_0, g_0) and any ψ_0 we can reconstruct the dynamics from a time series of ψ_0 but merely that we can do so for some (f, g) and ψ that can be chosen arbitrarily close to (f_0, g_0) and ψ_0. Now, observe that even if (f_0, g_0) and ψ_0 are of the form above for forced systems (i.e., with g_0 independent of x and ψ_0 independent of y), there is no reason why this should be true for the nearby (f, g) and ψ. Indeed, for typical (f, g) and ψ in the neighbourhood of (f_0, g_0) and ψ_0, g will be a function of x as well as y and ψ will depend on y.

As an example, suppose that we attempted to apply Takens Theorem to a periodically forced system, and for simplicity ignore perturbations in f and g. Then we would be able to conclude that arbitrarily close to our observation function $\psi_0(x, \theta) = \varphi_0(x)$ there was an observation function $\psi(x, \theta)$ for which reconstruction was possible, but unfortunately typically ψ would have nontrivial dependence on θ. In other words, *our measurements would be dependent on the phase of the forcing*. In most practical applications, this will not be the case: e.g., if we measure the displacement of a forced pendulum, or a voltage in a forced electronic oscillator, we do not expect our measurement apparatus to give explicitly time-dependent results.

Thus to summarize, forced systems and their observation functions are not typical in the space of all systems and all observation functions, and hence if we attempt to use Takens Theorem to justify our analysis of observed data, the conclusions that we arrive at are not the ones we would usually desire.

The aim of this paper is therefore to investigate to what extent Takens Theorem remains valid if we restrict ourselves to the space of forced systems and their observation functions. It turns out that, apart from minor restrictions on the forcing system g, we can in fact prove an exact analogue of Takens Theorem relevant to this case that allows us to reconstruct both f and g from a typical observable φ that is independent of y. We also prove a second version of this theorem that is more appropriate to situations where we already know the state y of the forcing system; this will typically be the case where we are responsible for the forcing ourselves, as in many laboratory experiments.

The one serious difficulty with these theorems is that in the case of periodically forced systems the forcing dynamics $\theta_{t+\tau} = \theta_t + \tau/T \pmod 1$ does not satisfy the conditions that we need to impose on g if t/T is rational with a "small" denominator. We can overcome this by weakening our notion of "reconstruction" in this case, but nevertheless we must stress that care needs to be taken in applying Takens Theorem to periodically forced systems where the sampling interval is rationally related to the forcing period. It would be interesting to see whether it is possible to prove a "full" Takens Theorem in this case.

The paper is organized as follows: In the next section, we give a precise statement of Takens Theorem, and discuss more fully its significance and its applications. In the following section we develop the formalism of forced systems, give precise statements of the two theorems we prove, present a number of examples that demonstrate the need for restrictions on the forcing system, and examine in detail the case of periodically forced differential equations. Takens' original proof consisted essentially of an easy local argument followed by a standard globalization procedure. Unfortunately, many technical details get in the way and obscure the fundamental concepts behind the proof. This would be even worse if we attempted this kind of approach in the forced case. Instead we base our proofs on Abraham's Parametric Transversality Theorem, which provides a powerful technique for gluing together local results and was in fact already used in this context by Aeyels [1981]. For the benefit of the reader unfamiliar with Abraham's Theorem, we outline the basic ideas behind transversality in Appendix A. To apply the Parametric Transversality Theorem, we need to regard spaces of functions between manifolds as infinite dimensional manifolds. There is a well developed theory of such function manifolds, and we gather together the results needed here in Appendix B. Appendix C contains a variety of technical calculations used throughout the paper. In Section 4, we present a self-contained proof of the standard Takens Theorem, using Abraham's Parametric Transversality Theorem. The reason for giving yet another proof of Takens Theorem is primarily to illustrate our approach to embedding theorems using transversality. The same ideas arise again in Sections 5 and 6, which contain proofs of the two versions of the forced Takens Theorem. However, given the added technical complexities involved in the forced theorems, we believe that it is helpful to the general reader to see these methods first presented in Section 4, in the simpler setting of the unforced case.

2. Takens Theorem

The behaviour of many real systems is believed to be at least approximately described by a deterministic finite-dimensional dynamical system. By this we mean that the state of the system is determined at any given instant by a point x lying on a m-dimensional manifold M that will often be some subset of $\mathbb{R}^k$. The time evolution of the system is given by a map $f^t\colon M \times \mathbb{R} \to M$ such that if the system is in state x_0 at time t_0 then it is in state $f^t(x_0)$ at time $t_0 + t$. In most applications, the map f^t will be obtained as the solution of some ordinary differential equation on M.

2.1. Observed Time Series

In many practical situations we do not have access to the state x and can merely observe some function $\varphi(x)$ of it. Here $\varphi\colon M \to \mathbb{R}$ is called the *measurement function*, and it corresponds to measuring some observable property of the system such as position or temperature. The evolution of this quantity with time is then given by $\varphi(f^t(x_0))$. In practice we can only observe this at discrete time intervals. For simplicity we assume that these intervals are all the same so that in fact we observe the sequence $\varphi_n = \varphi(x_n)$ for $n = 1, 2, \ldots$, where $x_n = f^{n\tau}(x_0)$ and $\tau > 0$ is the *sampling interval*. For obvious reasons $\{\varphi_n\}$ is called a *time series*. By rescaling time we may as well assume that $\tau = 1$, so that the sequence of states $\{x_n\}$ is given by the discrete dynamical system $x_{n+1} = f(x_n)$ where $f = f^1$ and f^n is just f composed n times.

The observed measurement φ_n is one-dimensional whilst x_n in general lies in some higher dimensional space. At first sight it might thus appear that φ_n contains relatively little information about the behaviour of x_n and that the fact that φ_n originates in the deterministic process f^n is of little use. However, Takens [1980] (see also Eckmann and Ruelle [1985], Sauer et al. [1991], Noakes [1991], and Huke [1993]) proved a remarkable theorem showing that for typical f and φ it is possible to reconstruct f up to some (unknown) smooth coordinate change.

2.2. Delay Embedding

More precisely, fix some d (called the *embedding dimension*) and define the *delay embedding map* $\Phi_{f,\varphi}\colon M \to \mathbb{R}^d$ by

$$\Phi_{f,\varphi}(x) = (\varphi(x), \varphi(f(x)), \ldots, \varphi(f^{d-1}(s)))^{\dagger}.$$

Suppose that M is compact, let $\mathcal{D}^r(M)$ be the set of $\mathcal{C}^r$ diffeomorphisms of M, and $\mathcal{C}^r(M, \mathbb{R})$ the set of observation functions on M, both endowed with the $\mathcal{C}^r$ topology. Recall that this is the topology of uniform convergence of a map and its derivatives up to r^{th} order; thus two functions are close in this topology if they and their derivatives are uniformly close (see, e.g., Hirsch [1976] for a precise definition).

Next we need to give a rigorous definition of genericity. We say that a property is *generic* in a function space $\mathcal{X}$ if it holds on a *residual* subset $\mathcal{A} \subset \mathcal{X}$, that is, on a subset that contains a countable intersection of open dense sets. Since $\mathcal{D}^r(M)$ and $\mathcal{C}^r(M, \mathbb{R})$ are complete metrizable spaces, such a set is itself dense by Baire's Theorem (e.g., Hirsch [1976]). This definition of generic is one of the most commonly used notions of "typical" in the field of dynamical systems (though not the only one; e.g., see Sauer et al. [1991]).

Finally, recall that a smooth map $\Psi\colon M \to N$ between manifolds M and N is called an *embedding* if it maps M diffeomorphically onto its image; if M is compact, this is equivalent to both Ψ and its derivative $T_x\Psi$ being injective (i.e., $1-1$) on the whole of M. Takens Embedding Theorem then states:

Theorem 2.1 (Takens [1980]). *Let M be a compact m dimensional manifold. Then if $d \geq 2m+1$, the set of (f, φ) for which the map $\Phi_{f,\varphi}$ is an embedding is open and dense in $\mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$ for $r \geq 1$.*

Note that Takens first proved this for $r \geq 2$, but it is relatively easy to extend this to $r = 1$ (see Huke [1993], or Section 4 below). A weaker version of this theorem was proved independently by Aeyels [1981], who considered the case where the sampling was not carried out at regular intervals, and showed that for generic choices of sampling times Φ was injective (and stated that it was also immersive). This is in fact much easier to prove, since one can ignore both periodic points of period less than d, and pairs of points (x, x') such that $x' = f^i(x)$; as we shall see below, these are precisely the two classes of points that cause us the most difficulty in the proof of Takens Theorem and its extensions. Nevertheless, it is clear that Aeyels' work contains all the fundamental ingredients needed to develop a dynamical-systems-based approach to time-series analysis.

It turns out to be possible to give a simple characterization of the set of f for which $\Phi_{f,\varphi}$ is an embedding. In particular, as Huke [1993] points out, Takens in fact deduces Theorem 2.1 from the following "unstated Takens Theorem":

Theorem 2.2. *Suppose that $f \in \mathcal{D}^r(M)$ has only a finite number of periodic orbits of period less than d, and the eigenvalues of each such periodic orbit are distinct. If $d \geq 2m+1$, then there is an open and dense set of $\varphi \in \mathcal{C}^r(M, \mathbb{R})$ for which $\Phi_{f,\varphi}$ is an embedding.*

Note that f satisfying the conditions of this theorem are open and dense in $\mathcal{D}^r(M)$. This is a simple generalization of the first part of the Kupka-Smale Theorem (e.g., Smale [1963]); a proof is also given in Section 4 below. The condition on the eigenvalues may seem a little strange, but is in fact almost a necessary condition for embedding. Thus:

Lemma 2.3. *Suppose that $f \in \mathcal{D}^r(M)$ has a fixed point $x \in M$ such that $T_x f$ has two linearly independent eigenvectors with the same eigenvalue. Then $\Phi_{f,\varphi}$ fails to be an immersion at x for all $\varphi \in \mathcal{C}^r(M, \mathbb{R})$.*

Proof. Let v_0 and v_1 be two linearly independent eigenvectors with eigenvalue λ. Then we can find constants $c_0, c_1 \neq 0$, such that $c_0 T_x\varphi(v_0) + c_1 T_x\varphi(v_1) = 0$. Let $u =$

$c_0 v_0 + c_1 v_1$. Then $T_x f^i(u) = \lambda^i u$, and hence $T_x(\varphi \circ f^i)(u) = \lambda^i T_x \varphi(u) = 0$. Thus, $T_x \Phi_{f,\varphi}(u) = 0$, and hence $T_x \Phi_{f,\varphi}$ is not $1-1$. □

More generally, if x is a periodic point of period $q < d$, then $\Phi_{f,\varphi}$ fails to be an immersion at x if there are more than q linearly independent eigenvectors of $T_x f^q$ with the same eigenvalue, and will be an immersion for an open dense set of φ if there are q or less such eigenvectors. One can also take account of nontrivial Jordan Normal Forms. It does not seem worthwhile, however, to include such detailed conditions in the statement of Takens Theorem, and the condition given in Theorem 2.2 represents a reasonable compromise between ease of presentation and the sharpest possible result.

2.3. Implications of Takens Theorem

Informally, Takens Theorem says that for typical f and φ, the image $\Phi(M)$ of M under $\Phi = \Phi_{f,\varphi}$ is completely equivalent to M itself, apart from the smooth invertible change of coordinates given by Φ. Furthermore, since Φ has a smooth inverse, we can define the map $F = \Phi \circ f \circ \Phi^{-1}$ on $\Phi(M)$. Then F is the same dynamical system as the original system given by f on M, but seen in the new coordinates given by the coordinate change Φ. In particular, all the coordinate-independent properties of F and f will be identical. This includes such features as the numbers and topological types of fixed points, periodic orbits, and other invariant sets as well as such geometric invariants as the eigenvalues of fixed and periodic points, and the correlation dimension and Liapunov exponents of corresponding invariant measures (e.g., see Eckmann and Ruelle [1985], Grassberger et al. [1992], Abarbanel et al. [1993], or Ott et al. [1994]). Thus, in particular, one can measure quantities such as the correlation dimension or the Liapunov exponents of F, and be certain that these are the same as those of the original (unknown) system f. This is important because these invariants of F can be estimated directly from the observed time series $\{\varphi_n\}$. This is because if we define the point $z_n \in \mathbb{R}^d$ by the delay coordinates $z_n = (\varphi_n, \varphi_{n+1}, \ldots, \varphi_{n+d-1})^\dagger$, then

$$\begin{aligned} z_n &= (\varphi(f^n(x_0)), \varphi(f^{n+1}(x_0)), \ldots, \varphi(f^{n+d-1}(x_0)))^\dagger \\ &= (\varphi(x_n), \varphi(f(x_n)), \ldots, \varphi(f^{d-1}(x_n)))^\dagger \\ &= \Phi(x_n). \end{aligned}$$

Thus z_n is in the image of Φ and so we can apply F to it:

$$\begin{aligned} F(z_n) &= \Phi \circ f \circ \Phi^{-1}(z_n) \\ &= \Phi \circ f \circ \Phi^{-1}(\Phi(x_n)) \\ &= \Phi \circ f(x_n) \\ &= \Phi(x_{n+1}) \\ &= z_{n+1}. \end{aligned}$$

Thus the dynamics of F on $\Phi(M)$ simply consists of sliding d successive elements of the time series $\{\varphi_n\}$ along by one time step, i.e.,

$$F(\varphi_n, \varphi_{n+1}, \ldots, \varphi_{n+d-1}) = (\varphi_{n+1}, \varphi_{n+2}, \ldots, \varphi_{n+d}).$$

As we noted above, this is completely equivalent to the original dynamics $x_{n+1} = f(x_n)$ up to the coordinate change Φ, and all the coordinate invariant properties of F and f are identical. Yet, whilst the original dynamics was unobservable, the dynamics in terms of F and z_n is expressible purely in terms of the observed time series $\{\varphi_n\}$. Thus if we have a sample of $\{\varphi_n\}$, we can deduce many of the properties of F, and hence of f. Thus to summarize, Takens Theorem allows us to reconstruct the unknown dynamical system that gave rise to the time series simply by constructing a new state space out of successive observations of the time series.

There are of course many practical obstacles that need to be surmounted if one is to use this in practice. Thus, for instance, we usually do not know m, and hence do not know what value of embedding dimension d we should use. Similarly the choice of sampling interval τ can significantly affect the performance of practical techniques. Such difficulties can however be overcome (e.g., see Abarbanel et al. [1993], or Ott et al. [1994]), and today the estimation of fractal dimensions or Liapunov exponents from time series is a fairly straightforward matter.

One can also go further, and attempt to forecast future values of the time series. In particular, consider the map F. Its first $d - 1$ components are trivial, since they consist of just shifting the argument of F by one time step. Denote the last component of F by $G\colon \Phi(M) \to \mathbb{R}$. Then $G(\varphi_n, \varphi_{n+1}, \ldots, \varphi_{n+d-1}) = \varphi_{n+d}$, or in other words, G simply predicts the time series $\{\varphi_n\}$ one time step ahead! Thus the time series $\{\varphi_n\}$ is completely deterministic and hence in principle entirely predictable (though if f and hence F is chaotic there is an upper limit on how far into the future we can predict in practice). Of course, we usually do not know G, but once again we can estimate it from a sample of the time series. Not only can this be useful in itself, but it also forms the basis of a variety of techniques for noise reduction, signal separation, control, and synchronization (again, see Abarbanel et al. [1993] or Ott et al. [1994]).

3. Forced Systems

In many applications, one encounters systems that are driven by some second system g. By far the best known case is that of periodic forcing. This arises in many laboratory experiments, where some kind of forcing is required to elicit nontrivial dynamical behaviour. Thus, for instance, a quick glance at a reprint collection such as Ott et al. [1994] reveals a large number of papers involving periodic forcing (e.g., Moon and Holmes [1979], Flepp et al. [1991], Sommerer et al. [1991], Papoff et al. [1992], Ditto et al. [1990], Hunt [1991], Gills et al. [1992], Shinbrot et al. [1992]). Indeed virtually all mechanical and electrical oscillators need to be driven if they are not simply going to rest at a trivial equilibrium, and the same can be said of lasers and related devices. At the other end of the applications spectrum, it would be surprising if for instance the population of a given species within some ecosystem was not sensitive to the annual seasonal cycle; a closely related example is given by the well-known fact that measles epidemics are affected by the scholastic year [Grenfell, 1992].

A simple generalization of periodic forcing is that of so-called quasi-periodic forcing, i.e., forcing by two periodic signals at incommensurate frequencies. This often appears to lead to very interesting dynamical behaviour (e.g., Romeiras et al. [1987]). An interesting example of more or less arbitrary forcing occurs in the context of synchronization (Pecora and Carroll [1990], [1991]), where a given system is used to drive a separate copy of a part of itself. This has potential applications to communications systems (e.g., Cuomo and Oppenheim [1993], Hayes et al. [1993]). Another situation where arbitrary driving dynamics can arise is in the adaptive prediction of chaotic time series (Stark [1993], Stark and Davies, [1994]) and more generally in the recursive filtering of such time series (e.g., Badii et al. [1988], Broomhead et al. [1992]).

3.1. Skew Products

If we want to study such forced systems using the techniques of nonlinear dynamics, we have to turn them into autonomous systems in the usual way by expanding the phase space and incorporating the forcing dynamics. As already described in the introduction, this leads to a dynamical system on $M \times N$ of the form

$$\begin{aligned} x_{i+1} &= f(x_i, y_i), \\ y_{i+1} &= g(y_i). \end{aligned}$$

Here x_i, as before, represents the state of the system we are interested in, and g represents the forcing dynamics. In the same way that in the unforced case f was required to be a diffeomorphism, we shall need the map $f_y\colon M \to M$ to be a diffeomorphism of M for every $y \in N$, where f_y is defined by $f_y(x) = f(x, y)$. This will be the case, for instance, if the underlying dynamics is in fact given by a differential equation (whose time τ map is f, see e.g., Example 3.1 below); this would seem to be a reasonable hypothesis in most of the examples described above. We shall employ the notation $\mathcal{D}^r(M \times N, M)$ for the set of such f in $\mathcal{C}^r(M \times N, M)$. We shall also need g to be a diffeomorphism of N, which in turn means that the pair (f, g) is a diffeomorphism of $M \times N$. In line with common usage, we shall call the pair (f, g) a skew product on $M \times N$.

Example 3.1. A particularly illustrative example arises in the case of periodically forced ordinary differential equations:

$$\frac{dx}{dt} = \Psi(x, t),$$

where $\Psi(x, t)$ is a time-dependent vector field that is periodic in time, so that $\Psi(x, t + T) = \Psi(x, t)$ for some period $T > 0$. Let f^s be the solution of the differential equation, so that $f^s(x, t)$ is the state of the system at time $t + s$, if the system was in state x at time t.

Suppose that we observe a measurement function $\varphi\colon M \to \mathbb{R}$ with a sampling interval $\tau > 0$. Then the i^{th} observation in our time series is given by $\varphi(f^{i\tau}(x, 0)) = \varphi(x_i)$ where x_i satisfies $x_{i+1} = f^t(x_i, t_i)$ with $t_{i+1} = t_i + \tau$. But $f^\tau(x_i, t_i) = f^\tau(x_i, t_i + T)$, and hence if we write $\theta_i = t_i/T$ we get the skew product on $M \times \mathbb{T}^1$, where $\mathbb{T}^1$ is the unit circle, given by

$$\begin{aligned} x_{i+1} &= f(x_i, t_i), \\ \theta_{i+1} &= \theta_i + \omega \qquad (\text{mod}\, 1), \end{aligned}$$

where $f(x,\theta) = f^\tau(x,\theta T)$ and $\omega = \tau/T$. In this case, therefore, g is simply a rigid rotation of the circle through the angle ω. If the sampling interval τ is incommensurate to the period T, then ω is irrational and the orbit of any θ is dense in $\mathbb{T}^1$, whilst if the sampling interval is rationally related to the forcing period, then ω is rational and every orbit of g is periodic. In many circumstances, particularly within the context of a laboratory experiment, one will of course sample at the forcing period, so that $\tau = T$ and g is the identity $g(\theta) = \theta$.

3.2. Embedding Skew Products

Given how common forced systems are, and the fact that in analysing time series produced by them it is normal to at least implicitly appeal to Takens Theorem, it is natural to ask what this theorem can tell us about delay embeddings of skew products. At first sight, there is no difficulty in applying the theorem to this case; after all (f, g) is a diffeomorphism of $M \times N$, and if both M and N are compact, then so is $M \times N$. We can thus conclude that "typically" the map $\Phi_{f,g,\psi}\colon M \times N \to \mathbb{R}^d$ is an embedding, where $\Phi_{f,g,\psi}$ is given by

$$\Phi_{f,g,\psi}(x, y) = (\psi(x, y), \psi((f, g)(x, y)), \psi((f, g)^2(x, y)), \ldots, \psi((f, g)^{d-1}(x, y)))^\dagger.$$

Unfortunately, as already indicated in the introduction, the notion of typical that is given by Takens Theorem to us here is not the one that we usually want in practical applications. More precisely, the theorem says that $\Phi_{f,g,\psi}$ is an embedding for an open dense subset of $\mathcal{D}^r(M\times N)\times\mathcal{C}^r(M\times N,\mathbb{R})$, whilst we want it to be an embedding for an open dense set in $\mathcal{D}^r(M \times N, M) \times \mathcal{D}^r(N) \times \mathcal{C}^r(M, \mathbb{R})$. Thus, since $\mathcal{C}^r(M, \mathbb{R})$ is not generic in $\mathcal{C}^r(M \times N, \mathbb{R})$, and $\mathcal{D}^r(M \times N, M) \times \mathcal{D}^r(N)$ is not generic in $\mathcal{D}^r(M \times N)$, we cannot conclude that typical skew products and typical functions on M lead to an embedding. As already mentioned in the introduction, the difference between the two statements is clearly highlighted in the case of periodically forced systems (as in Example 3.1): $\mathcal{C}^r(M,\mathbb{R})$ is the space of ordinary time-independent observation functions, whilst typical functions in $\mathcal{C}^r(M \times \mathbb{T}^1, \mathbb{R})$ have explicit time dependence (or more precisely, phase dependence).

We thus conclude that the existing versions of the Takens Theorem are not relevant to forced systems. The principal aim of this paper is therefore to prove versions of the theorem applicable to skew products. Since we shall only be concerned with observation functions of the form $\varphi\colon M \to \mathbb{R}$, we can write the delay map $\Phi_{f,g,\varphi}\colon M \times N \to \mathbb{R}^d$ as

$$\Phi_{f,g,\varphi}(x, y) = (\varphi(f^{(0)}(x, y)), \varphi(f^{(1)}(x, y)), \ldots, \varphi(f^{(d-1)}(x, y)))^\dagger,$$

where $f^{(i)}\colon M \times N \to M$ is given by $f^{(i+1)}(x, y) = f(f^{(i)}(x, y), g^i(y))$ with $f^{(0)}(x, y) = x$ and hence $f^{(1)}(x, y) = f(x, y)$. Ideally, we would like to conclude that if $d \geq 2(m + n) + 1$, then this is an embedding for generic f and generic φ, for any given g; in other words, we might hope that there is sufficient freedom in perturbing f and φ to allow us to fix the forcing system. Unfortunately, this is not the case: For instance, Example 3.3 in Section 3.5 below shows that if we take $N = \mathbb{T}^1$ and $g(\theta) = \theta$, then we can construct open sets of f and φ for which $\Phi_{f,g,\varphi}$ is not an embedding. Recall from Example 3.1 that this choice of N and g is precisely that which occurs for a periodically forced differential equation when the sampling interval is the same as the forcing period. The basic idea behind this example is quite simple: Since every point of N is a fixed point of g, it turns out to be easy to construct an open set of f for which (f, g) has a one-dimensional set of fixed points. At such points, every component of $\Phi_{f,g,\varphi}$ is identical, and so if φ has say a maximum near this set, $\Phi_{f,g,\varphi}$ will be not be $1-1$. Since details of this example are somewhat technical and would detract from the exposition here, they are postponed until Section 3.5 below.

A second, very simple, counterexample shows that we also need to exclude the case of zero-dimensional M. This should be no surprise, for if M consists of discrete points, and $f\colon M \times N \to M$ is continuous, then f, and hence $\Phi_{f,g,\varphi}$, must be independent of y.

Example 3.2. Take M to be a single point, i.e., $M = \{x\}$. Then of course there is only one map in $\mathcal{D}^r(M \times N, M)$, namely that given by $f(x, y) = x$. Thus $\Phi_{f,g,\varphi}(M \times N)$ is the single point $\{(c, c, \ldots, c)^\dagger\}$, where $c = \varphi(x)$. Hence, if N consists of more than one point, $\Phi_{f,g,\varphi}$ cannot be an embedding for any f, g, φ, let alone a residual set.

More generally, suppose that $M = \{x_1, \ldots, x_p\}$ consists of a finite number of points. Let f again be given by $f(x, y) = x$, for all $x \in M$, and note that any sufficiently small neighbourhood of f in $\mathcal{D}^r(M \times N, M)$ contains only f. But $\Phi_{f,g,\varphi}(x_i) = (c_i, c_i, \ldots, c_i)$ where $c_i = \varphi(x_i)$ and hence $\Phi_{f,g,\varphi}(M \times N)$ consists of precisely p points. Hence, if N consists of more than one point, $\Phi_{f,g,\varphi}$ is not an embedding for an open set of f, for any g and φ.

On the whole, this counterexample does not particularly concern us since the dynamics of zero-dimensional manifolds is not exceptionally interesting, and hence the failure of Takens Theorem in this case is not unduly worrying.

These two pathologies turn out to be the only obstructions to the genericity of embeddings for forced systems. We can exclude the first one by requiring the periodic orbits of short period of g to be isolated; and as it turns out, we also need nondegeneracy conditions on their eigenvalues. The condition we impose is the same as that on f in Theorem 2.2, but for technical reasons we require that it hold up to period $2d$. The precise version of Takens Theorem for skew products that we prove here is thus given by the following.

Theorem 3.1 (Forced Takens Theorem). *Let M and N be compact manifolds of dimension $m \geq 1$ and n, respectively. Suppose that the periodic orbits of period $< 2d$ of $g \in \mathcal{D}^r(N)$ are isolated and have distinct eigenvalues, where $d \geq 2(m+n)+1$. Then for $r \geq 1$, there exists an open and dense set of $(f, \varphi) \in \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$ for which the map $\Phi_{f,g,\varphi}$ is an embedding.*

The proof is given in Section 5, whilst Proposition 4.1 below shows that the set of g satisfying the conditions of this theorem is open and dense in $\mathcal{D}^r(N)$.

Given the discussion in Section 3.2, it would be interesting to see whether Theorem 3.1 generalizes to fully coupled systems, i.e., where g is not independent of x. Informally this would say that one can reconstruct coupled dynamical systems by observing a single subsystem. The appropriate conjecture would be that $\Phi_{f,g,\varphi}$ is an embedding for an open dense set of $((f, g), \varphi) \in \mathcal{D}^r(M \times N) \times \mathcal{C}^r(M, \mathbb{R})$. It seems likely that it should be possible to prove this using similar techniques to those developed here, and this issue will be addressed in a subsequent paper.

3.3. Bundle Embeddings

The Forced Takens Theorem above shows that we can typically reconstruct the joint dynamics of a forcing and a forced system from a time series of an observable of the forced system alone. It is applicable to situations where the forcing is unknown, which will be the case in many practical applications.

Sometimes, however, we may be in possession of independent knowledge of the state y of the forcing system. This will for instance be the case if we control the forcing, as in many laboratory experiments. If this is so, it seems that it should be unnecessary to have to reconstruct the forcing dynamics.

Thus, rather than requiring $\Phi_{f,g,\varphi}$ to embed $M \times N$, we might simply ask that it should embed each $M \times \{y\}$. More precisely, if we define $\Phi_{f,g,\varphi,y}\colon M \times \{y\} \to \mathbb{R}^d$ by $\Phi_{f,g,\varphi,y}(x) = \Phi_{f,g,\varphi}(x, y)$, then we would want $\Phi_{f,g,\varphi,y}$ to be an embedding for each $y \in N$. We shall call such a $\Phi_{f,g,\varphi,y}$ a *bundle* or *fibre embedding*.

An embedding of this kind is sufficient for instance if we have independent knowledge of $y_i = g^i(y)$ and want to predict the time series $\varphi(x_i)$. Thus, as in Section 2, let

$$\begin{aligned} z_i &= (\varphi(x_i), \varphi(f^{(1)}(x_i, y_i)), \ldots, \varphi(f^{(d-1)}(x_i, y_i)))^\dagger \\ &= \Phi_{f,g,\varphi}(x_i, y_i). \end{aligned}$$

Then if for some d both Φ_{f,g,φ,y_i} and $\Phi_{f,g,\varphi,y_{i+1}}$ are embeddings of $M \times \{y_i\}$ and $M \times \{y_{i+1}\}$, respectively, we have

$$z_{i+1} = \Phi_{f,g,\varphi,y_{i+1}} \circ f_{y_i} \circ (\Phi_{f,g,\varphi,y_i})^{-1}(z_i),$$

where $f_{y_i}(x) = f(x, y_i)$. Thus, if we let $G_{y_i}\colon \mathbb{R}^d \to \mathbb{R}$ be the last component of $\Phi_{f,g,\varphi,y_{i+1}} \circ f_{y_i} \circ (\Phi_{f,g,\varphi,y_i})^{-1}$, we have

$$\varphi_{i+d} = G_{y_i}(\varphi_i, \varphi_{i+1}, \ldots, \varphi_{i+d-1}),$$

where as usual $\varphi_i = \varphi(x_i)$. Then, in principle, given enough data we can estimate G_y as a function from $\mathbb{R}^d \times N$, and use this to predict the time series.

The main benefit of this approach is that we do not waste resources in trying to reconstruct the dynamics of g, which we already know. In particular, since we are only trying to embed an m dimensional manifold, we would expect it to be sufficient to take $d \geq 2m + 1$ rather than $d \geq 2(m + n) + 1$ as above. This in fact turns out to be the case. Since it is always desirable to work with the smallest d possible, this can be a significant advantage even if n is not particularly large.

The one drawback is that it is not possible to ensure that $\Phi_{f,g,\varphi,y}$ is an embedding for all $y \in N$. In particular, Example 3.4 below shows that it is possible to construct an open set of f, g, and φ for which $\Phi_{f,g,\varphi,y}$ fails to be an embedding at isolated y (depending on f, g, and φ). The best that we can do is show that it is an embedding for “typical” y. One would expect this to be sufficient in applications since it will allow us to construct G_y and make predictions for all “typical” y, in the same sense as one can use the standard Takens Theorem to construct G for “typical” f and φ.

Since N is finite-dimensional, there are two possible notions of “typical” y: either all y in an open dense set, or all y in a set of full Lebesgue measure. In fact the latter implies the former. In particular, the continuity of $\Phi_{f,g,\varphi,y}$ and the density of embeddings in $\mathcal{C}^r(M, \mathbb{R}^d)$ implies that the set of y for which $\Phi_{f,g,\varphi,y}$ is an embedding is open. Also, a set of full Lebesgue measure in N is necessarily dense since if it were not then its complement would contain an open set and hence have positive Lebesgue measure. The theorem we thus prove is as follows.

Theorem 3.2 (Bundle Delay Embedding Theorem). *Let M and N be compact manifolds of dimension $m \geq 1$ and n, respectively. Suppose that $d \geq 2m + 1$ and the closure of the set of periodic orbits of period $\leq d$ of $g \in \mathcal{D}^r(N)$ has zero Lebesgue measure in N. Then for $r \geq 1$, there exists a residual set of $(f, \varphi) \in \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$ such that for any (f, φ) in this set there is an open dense set of y of full Lebesgue measure such that $\Phi_{f,g,\varphi,y}$ is an embedding.*

The proof is given in Section 6. We can slightly weaken the condition on g to the requirement that the closure of the set of periodic orbits of period $\leq d$ is nowhere dense to give a residual set of (f, φ) for which $\Phi_{f,g,\varphi,y}$ is an embedding for an open dense set of y, but in this case we need $r \geq 2n$. It is an open question to what extent any conditions on g are necessary. In particular, one might hope that it might be possible for $\Phi_{f,g,\varphi,y}$ to typically be an embedding for all g. There appears to be no obvious counterexample to this conjecture, and in particular systems of the type considered in Example 3.1 fail to be $1 - 1$ only by mapping points with different y values to the same point, and are thus irrelevant here.

Finally, we remark that we have only been able to prove that Φ is a bundle embedding for a residual set of f and φ. This is a weaker result than in the standard Takens Theorem and Theorem 3.1 above, where Φ is an embedding for an open dense set of f and φ. In fact, in all three cases we only prove the existence of a residual set. However, for the latter two theorems we can then appeal to the fact that the map $(f, \varphi) \mapsto \Phi$ is continuous and that embeddings are open in $\mathcal{C}^r(M, \mathbb{R}^d)$ to deduce that the set of (f, φ) such that Φ is an embedding is also open. This argument fails for Theorem 3.2: If $\Phi_{f,g,\varphi,y}$ is an embedding, then certainly it is an embedding for a neighbourhood of (f, φ). However, the size of this neighbourhood depends on y, and hence there is no way of ensuring that there is a neighbourhood for which we get an embedding for typical y. This line of reasoning suggests one possible version of the theorem that gives embeddings for an open dense set of systems: Given any $\delta > 0$, there exists an open and dense set of (f, φ) for which $\Phi_{f,g,\varphi,y}$ is an embedding for a set of y of Lebesgue measure $1 - \delta$ (see Section 6.4).

It remains to be seen whether the distinction between open dense and residual is a significant one in applications. Certainly, many other results in nonlinear dynamics are restricted to a residual set of systems, and "residual" is widely accepted as a useful definition of "typical." Of course, if it proves unsatisfactory in this case, one can always increase the embedding dimension from $2m+1$ to $2(m+n)+1$, in which case Theorem 3.1 ensures that $\Phi_{f,g,\varphi,y}$ is an embedding for all y for an open dense set of (f, φ).

3.4. Periodically Forced Differential Equations

In this section we want to discuss the implications of the results stated above to periodically forced differential equations, continuing the analysis begun in Example 3.1. We also give detailed descriptions of the two examples (3.3 and 3.4) referred to in Section 3.2 and Section 3.4, respectively. Recall that the first of these shows that some restrictions on g are necessary in Theorem 3.1, and the second that we cannot expect to embed every fibre in Theorem 3.2.

As described in Example 3.1, periodically forced differential equations correspond to the case $N = \mathbb{T}^1$ and $g(\theta) = \theta + \tau/T$, where τ is the sampling interval and T is the forcing period. We thus see that when τ/T is a rational with denominator less than d, every point in $\mathbb{T}^1$ will be periodic (for g) with period less than d, and hence the hypotheses of Theorems 3.1 and 3.2 will unfortunately not be satisfied.

This is potentially a serious problem, since a very popular, and quite natural, choice is to sample at the forcing frequency, so that $\tau/T = 1$, $g = Id$, and every point in $\mathbb{T}^1$ is a fixed point. As Example 3.3 below shows, if the flow has a periodic orbit of period T that has a nontrivial projection onto M, then for an open set of φ, $\Phi_{f,g,\varphi}$ will fail to be $1-1$, and hence fail to be an embedding of $M \times \mathbb{T}^1$.

Fortunately, in most practical applications this is not a great drawback. This is because when one takes $\tau = T$, one is usually thinking in terms of reconstructing the so called stroboscopic map for the system, which is in effect the Poincaré map for the section $\theta = \theta_0$ (for some constant θ_0). Thus, one is trying to embed $M \times \{\theta_0\}$ for a single θ_0, rather than the whole of $M \times \mathbb{T}^1$. In such a case, we just have the single map f_{θ_0} to deal with (where $f_{\theta_0}(x) = f(x, \theta_0)$), and can justify the embedding procedure using the standard Takens Theorem. Thus for a residual set of f_{θ_0} and φ, the corresponding delay map (which is just $\Phi_{f,g,\varphi,\theta_0}$) will be an embedding.

The same idea can be used when τ/T is a rational with denominator less than d. Thus suppose that $\tau/T = p/q$, with $1 \leq q < d$. Given θ_0, define $\theta_1, \ldots, \theta_{q-1}$ by $\theta_{i+1} = g(\theta_i)$, and $\theta_0 = g(\theta_{q-1})$. Then the orbit of θ_0 is restricted to $N = \{\theta_0, \ldots, \theta_{q-1}\}$, which is a zero-dimensional manifold. Trivially, g has only a finite number of periodic points, and thus by Theorem 3.1 there is a dense open set of f such that $\Phi_{f,g,\varphi}$ embeds $M \times N = M \times \{\theta_0, \ldots, \theta_{q-1}\}$. Thus as long as we are satisfied with reconstructing the dynamics on $M \times N$ only, Theorem 3.1 is perfectly adequate.

Recently, however, there has been considerable interest in embedding the whole of $M \times \mathbb{T}^1$, for instance when it is desired to reconstruct the braid structure of the periodic orbits (e.g., Papoff et al. [1992], Tuffilaro et al. [1995] and the references therein). In such cases it is usual to highly oversample the time series, i.e., to take $\tau = T/q$, with q large, or τ/T irrational. By Theorem 3.1 the corresponding $\Phi_{f,g,\varphi}$ is generically an embedding.

Note also that Theorems 3.1 and 3.2 are phrased in terms of a residual set of f, whilst in practice we want genericity in terms of the original vector field Ψ. Thus, denote the set of periodically time-dependent vector fields by $\mathcal{C}^r(M \times \mathbb{T}^1, TM)$ (where we normalize the forcing period T to $T = 1$), and for $\Psi \in \mathcal{C}^r(M \times \mathbb{T}^1, TM)$ define the delay embedding map by

$$\Phi_{\Psi,\varphi}(x, \theta) = (\varphi(x), \varphi(f^{\tau}(x, \theta)), \ldots, \varphi(f^{\tau(d-1)}(x, \theta)))^{\dagger},$$

where f^t is the flow generated by Ψ, so that $f^t(x, \theta)$ is the state of the system at time $\theta + t$ (mod 1) if the state was x at time θ. It is then relatively straightforward to give the following analogue of Theorem 3.1.

Theorem 3.3. *Let M be a compact manifold of dimension $m \geq 1$. If $q\tau$ is not an integer for any $1 \leq q < d$, and $d \geq 2m+3$, then for $r \geq 1$ there exists an open and dense set of $(\Psi, \varphi) \in \mathcal{C}^r(M \times \mathbb{T}^1, TM) \times \mathcal{C}^r(M, \mathbb{R})$ for which $\Phi_{\Psi,\varphi}$ is an embedding of $M \times \mathbb{T}^1$. If $\tau = 1/q$ for some $1 \leq q < d$, and $d \geq 2m+1$, then for $r \geq 1$ there exists an open and dense set of $(\Psi, \varphi) \in \mathcal{C}^r(M \times \mathbb{T}^1, TM) \times \mathcal{C}^r(M, \mathbb{R})$ for which $\Phi_{\Psi,\varphi}$ is an embedding of $M \times \{\theta_0, \ldots, \theta_{q-1}\}$.*

We indicate the main modifications to the proof of Theorem 3.1 required to prove this in Section 5.11.

Finally, we give a detailed description of the two examples referred to above. The first constructs open sets of f and φ for which $\Phi_{f,g,\varphi}$ is not an embedding in the case $N = \mathbb{T}^1$ and $g(\theta) = \theta$, whilst the second gives an open set of f, g, and φ for which $\Phi_{f,g,\varphi,y}$ fails to be an embedding for every y. Of necessity, these examples are somewhat technical, and the casual reader may prefer to jump straight to Section 3.5.

Example 3.3. Take $N = \mathbb{T}^1$ and $g(\theta) = \theta$, and let M be any compact manifold. Let $\hat{f}$ be any diffeomorphism of M that has a hyperbolic fixed point x_0. Given $\hat{f}$, we can trivially construct a f_0: $M \times N \to M$ by $f_0(x, \theta) = \hat{f}(x)$. Then $(f_0, g) = (f_0, Id)$ is a skew product on $M \times N = M \times \mathbb{T}^1$, and (x_0, θ) is a fixed point for every $\theta \in \mathbb{T}^1$. Such fixed points are not hyperbolic, but the set $\{x_0\} \times \mathbb{T}^1$ is a normally hyperbolic invariant circle (e.g., Ruelle [1989]). Now, normally hyperbolic invariant manifolds are structurally stable under perturbations. Thus there is an open neighbourhood $\mathcal{U}$ of f_0 in $\mathcal{D}^2(M \times \mathbb{T}^1, M)$ such that for all f in $\mathcal{U}$ there is a (f, Id) invariant circle close to $\{x_0\} \times \mathbb{T}^1$. Furthermore this circle is the graph of some function χ: $\mathbb{T}^1 \to M$ that satisfies the equation $f(\chi(\theta), \theta) = \chi(\theta)$. Fix θ, and let $z = (\chi(\theta), \theta)$. If $v \in T_z(M \times \mathbb{T}^1)$, we can write it as $v = (v_1, v_2)$ with $v_1 \in T_{\chi(\theta)}M$ and $v_2 \in T_\theta \mathbb{T}^1$, and define the partial derivatives of f at z by $T_{1,z}f(v_1) = T_z f(v_1, 0)$ and $T_{2,z}f(v_2) = T_z f(0, v_2)$. Differentiating the equation $f(\chi(\theta), \theta) = \chi(\theta)$, we obtain $T_{1,z}f \circ T_\theta\chi + T_{2,z}f = T_\theta\chi$. Thus if $T_{2,z}f$ is injective (i.e., has rank 1) then $T_\theta\chi$ is injective, and hence χ is a local embedding in the neighbourhood of θ. Choose some f close to f_0 such that $f(x_0, \theta) = x_0$ and $T_{2,z}f$ is injective. Thus $\chi(\theta) = x_0$ and $T_\theta\chi$ is injective. Then there is an open neighbourhood V of θ in $\mathbb{T}^1$ such that $\chi(V)$ is a one-dimensional submanifold of M, with $x_0 \in \chi(V)$. Observe that for any point $x \in \chi(V)$ there exists some $\theta \in V$ such that $f(x, \theta) = x$, and hence $\Phi_{f,g,\varphi}(x, \theta) = (\varphi(x), \varphi(x), \ldots, \varphi(x))^\dagger$ for any φ: $M \to \mathbb{R}$.

Choose an observation function φ that has a nondegenerate maximum at x_0. Denote the level surfaces of φ by $L_{c,\varphi}$, thus $L_{c,\varphi} = \{x \in M\colon \varphi(x) = c\}$. Then for $c \neq \varphi(x_0)$ sufficiently close to $\varphi(x_0)$, the set $L_{c,\varphi}$ is an embedded $m-1$ dimensional sphere (enclosing x_0). For an open dense set of c, the curve $\chi(V)$ will intersect $L_{c,\varphi}$ transversally at two (or more) points. Call these x_1 and x_2; thus $\varphi(x_1) = \varphi(x_2)$. Also by the above there exists $\theta_1, \theta_2 \in V$ such that $f(x_1, \theta_1) = x_1$ and $f(x_2, \theta_2) = x_2$. Hence $\Phi_{f,g,\varphi}(x_1, \theta_1) = (\varphi(x_1), \varphi(x_1), \ldots, \varphi(x_1))^\dagger = (\varphi(x_2), \varphi(x_2), \ldots, \varphi(x_2))^\dagger = \Phi_{f,g,\varphi}(x_2, \theta_2)$, and so $\Phi_{f,g,\varphi}$ is not 1-1, and hence not an embedding. Now observe that the manifolds $\pi \circ \chi(V)$ and $L_{c,\varphi}$ depend smoothly on f and φ, and hence will intersect in at least two points for an open neighbourhood of f and φ. Thus $\Phi_{f',g,\varphi'}$ will not be an embedding for all f' and φ' in such an open neighbourhood.

Example 3.4. (I am very grateful to an anonymous referee for suggesting this example.) Again take $N = \mathbb{T}^1$ but now let $g(\theta) = \theta + \omega \pmod 1$ with ω arbitrary (so that in particular we can choose ω so that there is an open neighbourhood of g such that no map in this neighbourhood has a periodic orbit of period 3 or less). We also take $M = \mathbb{T}^1$ and $f(x, \theta) = x + \rho(\theta) \pmod 1$, for some $\rho\colon \mathbb{T}^1 \to \mathbb{R}$. Initially we shall take $\rho = 1/2$, independent of θ, and then perturb it slightly to give f some θ dependence. Now choose $\varphi\colon \mathbb{T}^1 \to \mathbb{R}$ such that $\varphi(0) = \varphi(1/2) = 0$, φ is positive and has a unique maximum on $[0, 1/2]$ at $x = 1/4$, $\varphi(1/4 + \delta) > \varphi(1/4 - \delta)$ for all $\delta \in (0, 1/4)$, and $\varphi(x) = -\varphi(1/2 - x)$ for all $x \in [1/2, 1]$. Let d be the minimum embedding dimension required by Theorem 3.2, so that $d = 3$, and let $\Phi_{f,g,\varphi,\theta}$ be the corresponding delay map. Since f is independent of θ, so is $\Phi_{f,g,\varphi,\theta}$. Furthermore, since $\Phi_{f,g,\varphi,\theta}(0) = \Phi_{f,g,\varphi,\theta}(1/2) = (0, 0, 0)^\dagger$, we see that $\Phi_{f,g,\varphi,\theta}$ fails to be 1-1 and hence an embedding for all $q \in \mathbb{T}^1$. This is not that surprising, since f is so "nongeneric"; however, we shall show that if we slightly perturb ρ, then there is an open neighbourhood of the resulting f (and of g and φ as above) such that the delay map still fails to embed $\mathbb{T}^1 \times \{\theta\}$ for at least one θ.

To see this, denote the first two components of $\Phi_{f,g,\varphi,\theta}$ by $\tilde{\Phi} = (\tilde{\Phi}_1, \tilde{\Phi}_2)\colon \mathbb{T}^1 \to \mathbb{R}^2$. Observe that $\tilde{\Phi}(0) = \tilde{\Phi}(1/2) = 0$. The conditions on φ imply that for $x \in (0, 1/4)$, we have $\tilde{\Phi}_1(x) > -\tilde{\Phi}_2(x) > 0$, and for $x \in (1/4, 1/2)$, we have $-\tilde{\Phi}_2(x) > \tilde{\Phi}_1(x) > 0$. The image of $[0, 1/2]$ under $\tilde{\Phi}$ is thus a loop from $(0, 0)$ lying in the fourth quadrant. The condition $\varphi(x) = -\varphi(1/2 - x)$ implies that the image of $[1/2, 1]$ is this loop rotated through π, and hence lying in the second quadrant. The whole image $\tilde{\Phi}(\mathbb{T}^1)$ is thus a "figure eight" with the two arms crossing transversally at the origin.

Because of this transversality and since $\Phi_{f,g,\varphi,\theta}$ depends smoothly on f, g, φ, and θ, the image of the first two components of $\Phi_{f,g,\varphi,\theta}$ will still be such a "figure eight" for all sufficiently close f, g, φ, and θ. Fix some θ_0 and choose $\rho\colon \mathbb{T}^1 \to \mathbb{R}$ sufficiently close to 1/2 for this to hold for the resulting f and so that $\rho(\theta_0) = 1/2$, ρ is increasing on an interval $[\theta_0 - \delta, \theta_0 + \delta]$ for some $\delta > 0$ and $\rho(\theta) = \rho(\theta + \omega) = \rho(\theta + 2\omega)$ for $\theta \in [\theta_0 - \delta, \theta_0 + \delta]$. This last condition can be satisfied since we have assumed that g has no periodic orbits of period 3 or less, and hence the intervals $[\theta_0 - \delta, \theta_0 + \delta]$, $[\theta_0 + \omega - \delta, \theta_0 + \omega + \delta]$, and $[\theta_0 + 2\omega - \delta, \theta_0 + 2\omega + \delta]$ can be chosen distinct.

Observe that for any $\theta \in [\theta_0 - \delta, \theta_0 + \delta]$ we have $\Phi_{f,g,\varphi,\theta}(x) = (\varphi(x), \varphi(x + \rho(\theta)), \varphi(x + 2\rho(\theta)))^\dagger$, with $\rho(\theta)$ close to 1/2, $\rho(\theta) < 1/2$ for $\theta \in (\theta_0 - \delta, \theta_0)$, and $\rho(\theta) > 1/2$ for $\theta \in (\theta_0, \theta_{0+\delta}]$. Denote the first two components of the $\Phi_{f,g,\varphi,\theta}$ by $\tilde{\Phi}_{\rho,\theta}$. Consider first the case $\theta \in [\theta_0 - \delta, \theta_0)$. By the above, the image of $\tilde{\Phi}_{\rho,\theta}$ is a "figure eight". It is easy to see that the self-intersection cannot occur in the region $\varphi(x) \geq 0$. This is because the self-intersection must occur close to the origin but for $x \in (0, 1/4)$, we have $\varphi(x + \rho(\theta)) > \varphi(x + 1/2) > -\varphi(x)$ and for $x \in (3/4 - \rho(\theta), 1/2)$, we have $\varphi(x + \rho(\theta)) < \varphi(x + 1/2) < -\varphi(x)$. Hence the self-intersection must occur in the second quadrant and in particular correspond to $x_1 \in (-\varepsilon, 0)$ and $x_2 \in (1/2, 1/2 + \varepsilon)$ for some small $\varepsilon > 0$. Considering the third component of $\Phi_{f,g,\varphi,\theta}$, we then have $\varphi(x_1 + 2\rho(\theta)) < 0$ and $\varphi(x_2 + 2\rho(\theta)) > 0$. Turning now to $\varphi \in (\varphi_0, \varphi_0 + \delta]$, the same argument shows that the intersection occurs in the fourth quadrant, so that now $x_1 \in (0, \varepsilon)$ and $x_2 \in (1/2 - \varepsilon, 1/2)$, and hence $\varphi(x_1 + 2\rho(\theta)) > 0$ and $\varphi(x_2 + 2\rho(\theta)) < 0$. We thus see that as θ passes through θ_0, $\varphi(x_1 + 2\rho(\theta)) - \varphi(x_2 + 2\rho(\theta))$ undergoes a change of sign. Since $\varphi(x_1 + 2\rho(\theta)) - \varphi(x_2 + 2\rho(\theta))$ depends continuously on θ, it must pass through 0, which corresponds to $\Phi_{f,g,\varphi,\theta}(x_1) = \Phi_{f,g,\varphi,\theta}(x_2)$; by the above, this of course takes place at θ_0.

Now observe that x_1 and x_2 depend continuously on f, g, and φ. Hence for a sufficiently small open neighbourhood of f, g, and φ, we can find θ_1 and θ_2 such that $\varphi(x_1 + 2\rho(\theta_1)) - \varphi(x_2 + 2\rho(\theta_1)) < 0$ and $\varphi(x_1 + 2\rho(\theta_2)) - \varphi(x_2 + 2\rho(\theta_2)) > 0$, and hence a $\theta \in (\theta_1, \theta_2)$ such that $\varphi(x_1 + 2\rho(\theta_1)) = \varphi(x_2 + 2\rho(\theta_1))$, or in other words $\Phi_{f,g,\varphi,\theta}(x_1) = \Phi_{f,g,\varphi,\theta}(x_2)$ Hence, as claimed, we have an open neighbourhood of f, g, and φ such that $\Phi_{f,g,\varphi,\theta}$ fails to be 1-1 for at least one θ.

3.5. Stochastic Forcing

In many applications the assumption that the forcing is generated by a finite-dimensional deterministic system is not a reasonable one, and it would be extremely useful to develop a framework for reconstructing systems driven by far more general processes. Particularly examples include stochastic dynamical systems (which we think of as deterministic systems driven by some stochastic process), input-output systems (as considered for instance by Casdagli [1992]) and irregularly sampled time series. Since it is possible to extend Theorem 3.2 to cover this case, we give a brief overview of this approach. Full details will be contained in a sequel to this paper, jointly with D. Broomhead, M. Davies, and J. Huke.

A standard approach to modelling such systems is through the use of shift spaces. Thus let X be some topological space, and define $\Sigma = X^{\mathbb{Z}}$ to be the space of bi-infinite sequences of elements in X with the product topology. Let $\sigma\colon \Sigma \to \Sigma$ be the standard shift map; thus $[\sigma(\omega)]_i = [\omega]_{i-1}$, where $[\omega]_i$ is the i^{th} component of ω. Then if $f \in \mathcal{D}^r(M \times X, M)$, we get the skew product system,

$$
\begin{aligned}
x &\mapsto f(x, \omega_0), \\
\omega &\mapsto \sigma(\omega).
\end{aligned}
$$

Since the space Σ contains all possible sequences of elements in X, this gives us a very general model of systems driven by arbitrary sequences. Observe that one could more generally take $f \in \mathcal{D}^r(M \times \Sigma, M)$ and replace the x dynamics by $x \mapsto f(x, \omega)$. If f only depends on a finite number of components of ω, then such a generalization is straightforward to incorporate, whilst if it depends on an infinite number it is very unclear under what conditions reconstruction would still be possible.

So far we have considered ω as simply an arbitrary sequence. If, in addition, we have a probability measure μ on X, then the corresponding product measure μ_Σ on Σ is invariant under σ, giving rise to an interpretation of the above dynamics as a stochastic process. If we define the map $f_{\omega_i}: : M \to M$ by $f_{\omega_i}(x) = f(x, \omega_i)$, then one can think of such a process as applying a different map f_{ω_i} at each time step, with ω_i chosen randomly with respect to μ at each time step. One often takes $X = \mathcal{D}^r(M)$ here (so that $f_{\omega_i} = \omega_i(x)$), in which case the resulting system is called a *random diffeomorphism* (e.g. Kifer [1988]). Furthermore, when μ consists of a finite number of discrete atoms, one obtains an *iterated function systems* in the sense of Barnsley [1988] (see also Norman [1968]).

Whilst these kinds of systems do not encompass all possible cases of stochastic dynamics, they include a sufficiently rich class of systems to be both useful and interesting. It is also possible to consider more general shift invariant measures on Σ than just those given as product measures of a measure on X (e.g., corresponding to Markov processes).

Example 3.5. As a simple example, the reader might find it helpful to consider the case of an irregularly sampled periodic orbit of a differential equation. Restrict the dynamics to the periodic orbit and take our coordinate on the orbit to be just the phase $\theta \in \mathbb{T}^1$, so that the dynamics is just given by $\dot{\theta} = 1$. Suppose we observe the system at a sequence of times $\{t_i\colon i \in \mathbb{Z}\}$ and let $\omega_i = t_i - t_{i+1} \pmod 1$. Then, within the above framework we have $M = X = \mathbb{T}^1$ and $f_{\omega_i}(\theta) = \theta + \omega_i \pmod 1$. This is just a rotation of $\mathbb{T}^1$ through the angle ω_i and corresponds to the time ω_i map of the flow $\dot{\theta} = 1$. The dynamics of θ is therefore given by $\theta_{i+1} = \theta_i + [\sigma^i(\omega)]_0 \pmod 1 = \theta_i + \omega_i \pmod 1$. If one then assumes that say the ω_i are uniformly independently distributed on $\mathbb{T}^1$, this corresponds to taking μ as the Lebesgue (or more correctly, Haar) measure on $\mathbb{T}^1$.

The natural concept of equivalence for two skew product systems f and f' (over the same Σ) is that of a bundle conjugacy, that is, a map $h\colon M \times \Sigma \to M$ such that h_ω is a homeomorphism for each ω (where as usual $h_\omega(x) = h(x, \omega)$) and such that (h, Id) conjugates the skew dynamics of (f, Id) with the dynamics of (f', Id), i.e.,

$$(f, \sigma) \circ (h, Id) = (h, Id) \circ (f', \sigma).$$

This can be seen to be equivalent to

$$f_\omega \circ h_\omega = h_{\sigma(\omega)} \circ f'_\omega,$$

and is analogous to the concept of bundle embedding in Section 3.3. Ideally we would like $f_\omega \circ h_\omega = h_{\sigma(\omega)} \circ f'_\omega$ to hold for every $\omega \in \Sigma$, but in general as in Section 3.3 (recall Example 3.4) this is too ambitious, and we have to be content with a conjugacy only for "typical" ω, i.e., for generic ω in the topological setting, or for almost every ω with respect to μ_Σ in the measure theoretic context. Note that even when, as in our case, f depends only on ω_0, we still allow h to depend on the other components of ω.

Now suppose that we observe the dynamics using some function $\varphi\colon M \to \mathbb{R}$. If we define $f_{\omega_i \ldots \omega_0} = f_{\omega_i} \circ \cdots \circ f_{\omega_0}$, where, as above, $f_{\omega_i}(x) = f(x, \omega_i)$, then the usual delay embedding can then be written

$$\Phi_{f,\varphi,\omega}(x) = (\varphi(x), \varphi(f_{\omega_0}(x)), \ldots, \varphi(f_{\omega_{d-2}\ldots\omega_0}(x))).$$

By analogy to Theorem 3.2 above, if we wish to reconstruct the dynamics of our system using delay coordinates, then it is reasonable to require $\Phi_{f,\varphi}$ to be a bundle embedding, i.e., $\Phi_{f,\varphi,\omega}$ should be an embedding (in the usual sense) for "typical" ω. Recall that in Theorem 3.2 we made no attempt to reconstruct the forcing dynamics g, and as a consequence, the embedding dimension d was independent of the dimension n of the forcing system. Since the only difference between the setting here, and that of Theorem 3.2, is that we have replaced the finite-dimensional dynamics g by the infinite-dimensional shift map σ, it is not unreasonable to expect that we should still get a bundle embedding here. In fact, it turns out that as long as we take $X = N$, a finite-dimensional compact manifold, it is straightforward to modify Theorem 3.2 to give the following.

Theorem 3.4. *Let M and N be compact manifolds of dimension $m \geq 1$ and n, respectively. Suppose that $d \geq 2m + 1$. Then for $r \geq 1$, there exists an open dense set of $(f, \varphi) \in \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$ such that for any (f, φ) in this set there is an open dense set of ω in Σ such that $\Phi_{f,\varphi,\omega}$ is an embedding. If in addition μ is a measure on N that is absolutely continuous with respect to Lebesgue measure, we can ensure that $\Phi_{f,\varphi,\omega}$ is an embedding for μ_Σ-almost all ω.*

A proof of this theorem will be given in a subsequent paper, written jointly with D. Broomhead, M. Davies, and J. Huke. This also includes generalizations to cover the case of noise on the observation function φ.

We can think of the systems covered by this theorem as ones whose dynamics is given by a finite parameter family of diffeomorphisms on M, with the parameters being an *i.i.d.* random process with an absolutely continuous distribution. This would seem to be a sufficiently rich class of systems to encompass many applications to real systems, and the above theorem thus represents the first tentative step towards developing a theory of embedding nonlinear stochastic dynamical systems.

To conclude this section, we discuss the interpretation of Theorem 3.4 in the context of delay reconstruction of time series; this is virtually identical to that for Theorem 3.2 given in Section 3.3. Thus, suppose that w is such that $\Phi_{f,\varphi,\omega}$ and $\Phi_{f,\varphi,\sigma(\omega)}$ are both embeddings of M. Then the map $F_\omega = \Phi_{f,\varphi,\sigma(\omega)} \circ f_{\omega_0} \circ (\Phi_{f,\varphi,\omega})^{-1}$ is well defined and is a diffeomorphism between $\Phi_{f,\varphi,\omega}(M) \subset \mathbb{R}^d$ and $\Phi_{f,\varphi,\sigma(\omega)}(M) \subset \mathbb{R}^d$. Let $(x_i, \sigma^i(\omega))$ be an orbit of (f, σ), so that $x_{i+1} = f(x_i, \omega_i)$, and as usual define $\varphi_i = \varphi(x_i)$ and $z_i = (\varphi_i, \varphi_{i+1}, \ldots, \varphi_{i+d-1})$. Then $z_i = \Phi_{f,\varphi,\sigma^i(\omega)}(x_i)$, and hence if $\Phi_{f,\varphi,\sigma^i(\omega)}$ and $\Phi_{f,\varphi,\sigma^{i+1}(\omega)}$ are both embeddings, we have

$$\begin{aligned} z_{i+1} &= \Phi_{f,\varphi,\sigma^{i+1}(\omega)}(x_{i+1}), \\ &= \Phi_{f,\varphi,\sigma^{i+1}(\omega)}(f_{\omega_i}(x_i)), \\ &= \Phi_{f,\varphi,\sigma^i(\sigma(\omega))}(f_{\omega_i}((\Phi_{f,\varphi,\sigma^i(\omega)})^{-1}(z_i))), \\ &= F_{\sigma^i(\omega)}(z_i). \end{aligned}$$

Therefore in exact analogy to the standard Takens framework, $F_{\sigma^i(\omega)}$ is just the map that shifts a block of the time series forward by one time step, and hence $(\varphi_i, \varphi_{i+1}, \ldots, \varphi_{i+d-1}) \mapsto (\varphi_{i+1}, \varphi_{i+2}, \ldots, \varphi_{i+d})$ is bundle conjugate to our original dynamics f_{ω_i}. Note however, that whereas f_{ω_i} only depends on $\omega_i = \sigma^i(\omega)_0$, the map $F_{\sigma^i(\omega)}$ depends on $\omega_i, \omega_{i+1}, \ldots, \omega_{i+d-1}$. As usual, the first $d-1$ components of F_ω are trivial. If we denote the last component by $G_\omega\colon \Phi_{f,\varphi,\omega}(M) \to \mathbb{R}$ then

$$\varphi_{i+d} = G_{\sigma^i(\omega)}(\varphi_i, \varphi_{i+1}, \ldots, \varphi_{i+d-1}).$$

If we write out the dependence on $\sigma^i(\omega)$ explicitly, we get

$$\varphi_{i+d} = G(\varphi_i, \varphi_{i+1}, \ldots, \varphi_{i+d-1}, \omega_i, \omega_{i+1}, \ldots, \omega_{i+d-1}).$$

In the case of one-dimensional ω_i, the existence of such a function was conjectured by Casdagli [1992]. From another point of view, processes that satisfy this equation (again for a one-dimensional ω_i) are well known in signal processing under the name of Nonlinear Auto-Regressive Moving Average (NARMA) models (e.g., Billings et al.

[1988]). Note however, that in the case of NARMA processes, one usually assumes that G depends smoothly on all its arguments, whilst we can say nothing about the regularity of its dependence on $\omega_i, \omega_{i+1}, \ldots, \omega_{i+d-1}$; indeed, G is only defined for almost every $\omega_i, \omega_{i+1}, \ldots, \omega_{i+d-1}$ with respect to the product measure μ_d on N^d.

4. Proof of the Standard Takens Theorem

Takens Embedding Theorem can be thought of as an extension of the Whitney Embedding Theorem, which says that the set of embeddings of a compact manifold M in $\mathbb{R}^d$ is open and dense in $\mathcal{C}^r(M, \mathbb{R}^d)$ if $d \geq 2m + 1$. Unfortunately, maps of the form $\Phi_{f,\varphi}$ form a very special subset of $\mathcal{C}^r(M, \mathbb{R}^d)$, and hence we cannot deduce Takens Theorem directly from the Whitney Theorem. However, since the map $(f, \varphi) \mapsto \Phi_{f,\varphi}$ is continuous (by Corollary C.3, in Appendix C below), Whitney's Theorem does at least immediately imply that the set of (f, φ) such that $\Phi f, \varphi$ is an embedding is open in $\mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$. The main task in proving Takens Theorem is thus to demonstrate that there is sufficient independence in the components of $\Phi_{f,\varphi}$ to ensure that the set of (f, φ) giving an embedding is dense in $\mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$.

This is completely straightforward at those points $x \in M$ that do not lie on a periodic orbit of f of period less than or equal to d. Then the points $x, f(x), \ldots, f^{d-1}(x)$ are distinct, and hence we can perturb φ independently in the neighbourhood of each of these points (intuitively this is obvious; a rigorous argument can be found in Appendix C.3 below). Thus, informally, given any $\Psi\colon M \to \mathbb{R}^d$, we can find a $\varphi\colon M \to \mathbb{R}$ such that in a neighbourhood U of x the map $\Phi_{f,\varphi}$ agrees with Ψ. Since the map $\varphi \mapsto \Phi_{f,\varphi}$ is smooth, this means that the set of φ for which $\Phi_{f,\varphi}$ is an embedding restricted to U is dense by Whitney's Theorem. We then need to piece together these local embeddings into a global one. At this stage a further problem arises, namely to show that $\Phi_{f,\varphi}(x) \neq \Phi_{f,\varphi}(x')$ for points x and x' such that $x = f^k(x')$ for some $k \neq 0$ such that $-d \leq k \leq d$. The difficulty with such points is that perturbations in f and φ do not affect $\Phi_{f,\varphi}(x)$ and $\Phi_{f,\varphi}(x')$ independently. The vast part of the proof of Takens Theorem is taken in overcoming this, and in dealing with the short periodic orbits of f. As we shall see, exactly the same issues will arise in the proofs of the forced theorems in the next two sections.

4.1. Main Proof

Since M is assumed compact, to show that $\varphi_{f,\varphi}$ is an embedding it is sufficient to show that it is both injective and immersive. As already mentioned above, since the map $(f, \varphi) \mapsto \Phi_{f,\varphi}$ is continuous, the set of (f, φ) such that $\Phi_{f,\varphi}$ is an embedding is open by Whitney's Theorem, and it remains to show that it is dense (we shall in fact demonstrate that it is residual). Finally, since $\mathcal{C}^r(M, N)$ is dense in $\mathcal{C}^{r'}(M, N)$ for all $r > r'$, we only need to prove the theorem for all sufficiently large r; it turns out that we shall need at least $r = 3$.

The first step is to construct a set of (f, φ) such that $\Phi_{f,\varphi}$ is an embedding on the set of short periodic orbits of f. For convenience, we shall deal with all orbits of period less than $2d$. This differs from the approach of Takens [1980] and Huke [1993] who only consider periods up to d. The advantage of the present approach is that it avoids having to deal separately with points x and x' such that $x = f^k(x')$ and x, x' periodic with period q such that $d \leq q < 2d$. Furthermore, when we come to the proof of the forced theorem in Section 5, there seems to be no way of avoiding taking all periods up to $2d$. Since the present section is intended to provide an introduction to the techniques used to prove the forced version of the theorem, it seems preferable to use the same condition here.

We construct the required set of (f, φ) by restricting to those f that only have a finite number of periodic orbits of period less than $2d$. It is then a relatively simple matter to perturb φ in the neighbourhood of each such orbit to ensure that $\Phi_{f,\varphi}$ is 1-1 on the set of such orbits. Showing that $\Phi_{f,\varphi}$ is immersive takes somewhat more effort; the argument is motivated by the theory of Vandermode determinants, though we do not use this explicitly.

We proceed by defining

$$\mathcal{D}^r = \{f \in \mathcal{D}^r(M)\colon \text{ all periodic orbits of } f \text{ of period } q < 2d \text{ are isolated and hyperbolic and each has distinct eigenvalues}\}.$$

Note that we do not actually need the hyperbolicity of the periodic orbits in the proof of Takens Theorem, but have no independent way of proving that the set of f such that periodic orbits of a given period are isolated is dense. Furthermore, by the Hartman-Großman Theorem (e.g., [Irwin, 1980], [Ruelle, 1989]), a hyperbolic periodic orbit is isolated from periodic orbits of any fixed period, and hence the requirement that the short periodic orbits are isolated in the definition of $\mathcal{D}^r$ is superfluous. We include it in order to emphasize its significance in the proof of Takens Theorem. Also note that we shall not need the condition that orbits of period q for $d \leq q < 2d$ have distinct eigenvalues, and as discussed already in Section 2.2, even for $q < d$ this condition is slightly stronger than necessary. However, for clarity of presentation we prefer to define $\mathcal{D}^r$ as above.

The fact that the set of f such that the periodic orbits of a given period are hyperbolic forms the first part of the Kupka-Smale Theorem (e.g., Smale [1963]). It is then a simple matter to perturb f further to ensure that the eigenvalues of each such orbit are distinct, thereby giving

Proposition 4.1. *$\mathcal{D}^r$ is open and dense in $\mathcal{D}^r(M)$ for $r \geq 2$.*

We give a direct proof of this in Section 4.2 below. Given any $f \in \mathcal{D}^r$, let $\mathcal{P}_f$ be the set of periodic points of f of period less than $2d$. We now aim to construct an open and dense set $\mathcal{A}_f \subset \mathcal{C}^r(M, \mathbb{R})$ such that $\Phi_{f,\varphi}$ is 1-1 and immersive (i.e., $T\Phi_{f,\varphi}$ is 1-1) on $\mathcal{P}_f$. Since φ is the first component of $\Phi_{f,\varphi}$, the injectivity of $\Phi_{f,\varphi}$ follows immediately from

Proposition 4.2. *The set of φ such that φ is 1-1 on $\mathcal{P}_f$ is open and dense in $\mathcal{C}^r(M, \mathbb{R})$.*

Since $\mathcal{P}_f$ consists of a finite number of points, this is intuitively obvious, though a rigorous proof is given below in Section 4.2, where we also prove

Proposition 4.3. *The set of φ such that $T_x\Phi_{f,\varphi}$ has rank m for all $x \in \mathcal{P}_f$ is open and dense in $\mathcal{C}^r(M, \mathbb{R})$.*

Since $\Phi_{f,\varphi}$ is an immersion at x if and only if $T_x\Phi_{f,\varphi}$ has rank m, this shows that the set $\mathcal{A}_f$ of φ such that $\Phi_{f,\varphi}$ is an embedding on $\mathcal{P}_f$ is open and dense in $\mathcal{C}^r(M, \mathbb{R})$. One minor issue then remains: For any $f \in \mathcal{D}$ we have shown that $\mathcal{A}_f$ is open and dense in $\mathcal{C}^r(M, \mathbb{R})$, but in the statement of Takens Theorem we use an open and dense subset of $\mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$. We thus require

Lemma 4.4. *The set*

$$\mathcal{E}^r = \{(f, \varphi)\colon\ f \in \mathcal{D}^r,\ \varphi \in \mathcal{A}_f\}$$

is open and dense in $\mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$.

The proof is again given in Section 4.2 below.

From now on we shall restrict ourselves to $(f, \varphi) \in \mathcal{E}^r$. We shall show that $\Phi_{f,\varphi}$ is an embedding for an open dense subset of $\mathcal{E}^r$ by considering two maps ρ and σ from $\mathcal{E}^r$ into appropriate spaces of maps between M and $\mathbb{R}^d$, and their tangent bundles. The first of these maps $\rho\colon\ \mathcal{E}^r \to \mathcal{C}^1(\tilde{T}M, T\mathbb{R}^d)$ is defined by

$$\rho(f, \varphi) = \tilde{T}\Phi_{f,\varphi},$$

where $\tilde{T}M = \{v \in TM\colon \|v\| = 1\}$ is the unit tangent bundle of M, and $\tilde{T}\Phi_{f,\varphi}$ is just the restriction to $\tilde{T}M$ of the tangent bundle map $T\Phi_{f,\varphi}\colon :\ TM \to T\mathbb{R}^d$. Thus $ev_\rho(f, \varphi, v) = T_x\Phi_{f,\varphi}(v)$. Let L be the 0 section in $T\mathbb{R}^d$, i.e., $L = \{0_y \in T\mathbb{R}^d\}$, where 0_y is the zero in $T_y\mathbb{R}^d$. Note that $\Phi_{f,\varphi}$ is an immersion if and only if the image of $\tilde{T}\Phi_{f,\varphi}$ does not intersect L. We prove the following in Section 4.3 below.

Proposition 4.5. *The map ev_ρ is transversal to L.*

Now, by Corollary C.5, ev_ρ is $\mathcal{C}^1$ if $r \geq 3$, and if $d > 2m - 2$, then $1 > \dim \tilde{T}M - \operatorname{codim} L$. We may thus apply the Parametric Transversality Theorem to ρ. This allows us to conclude that the set of (f, φ) such that $\rho(f, \varphi)$ is transversal to L is open and dense in $\mathcal{E}^r$. But for any $v \in \tilde{T}M$, the dimension of $T_v(\tilde{T}M)$ is $2m - 1$, and hence the dimension of its image $T_v(\tilde{T}\Phi_{f,\varphi})(T_v(\tilde{T}M))$ is at most $2m-1$, whilst the dimension of T_uL is d for any $u \in L$. Hence if $\tilde{T}\Phi_{f,\varphi}(v) \in L$, $T_v(\tilde{T}\Phi_{f,\varphi})(T_v(\tilde{T}M)) + T_uL$ has dimension at most $2m - 1 + d$. If $d \geq 2m$, this is strictly less than $2d$, which is the dimension of $T_u(T\mathbb{R}^d)$, and hence $T_v(\tilde{T}\Phi_{f,\varphi})(T_v(\tilde{T}M)) + T_uL$ cannot possibly span $T_u(T\mathbb{R}^d)$. Thus if $d \geq 2m$, the only way that $\tilde{T}\Phi_{f,\varphi}$ can be transversal to L is if its image does not intersect L, i.e., if $\Phi_{f,\varphi}$ is an immersion. We have thus shown that the set of (f, φ) such that $\Phi_{f,\varphi}$ is an immersion is open and dense in $\mathcal{E}^r$.

Let us now turn to demonstrating the injectivity of $\Phi_{f,\varphi}$. For this we consider the map $\sigma\colon\ \mathcal{E}^r \to \mathcal{C}^1(M \times M\backslash\Delta, \mathbb{R}^d \times \mathbb{R}^d)$ defined by

$$\sigma(f, \varphi)(x, x') = (\Phi_{f,\varphi}(x), \Phi_{f,\varphi}(x')),$$

where Δ is the diagonal in $M \times M$. Let $\hat{\Delta} = \{(z, z)\colon z \in \mathbb{R}^d\}$ be the diagonal in $\mathbb{R}^d \times \mathbb{R}^d$, and observe that $\Phi_{f,\varphi}$ is injective if and only if its image does not intersect $\hat{\Delta}$. In Section 4.4 we shall show

Proposition 4.6. *The map ev_σ is transversal to $\hat{\Delta}$.*

By Corollary C.3, ev_σ is $\mathcal{C}^1$ for $r \geq 2$, and $\dim M \times M \backslash \Delta - \operatorname{codim} \hat{\Delta} = 2m - d < 1$, if $d > 2m - 1$. We may thus apply the Parametric Transversality Theorem to deduce that the set of (f, φ) such that $\sigma(f, \varphi)$ is transversal to $\hat{\Delta}$ is residual in $\mathcal{E}^r$.

But now, just as above, we count dimensions: The dimension of $T_{x,x'}(M \times M\backslash\Delta)$ is $2m$, and hence the dimension of $T_{x,x'}(\sigma(f, \varphi))(T_{x,x'}(M \times M\backslash\Delta))$ is less than or equal to $2m$. The dimension of $T_{z,z}\hat{\Delta}$ is d and thus if $2m + d < 2d$, transversality of $\sigma(f, \varphi)$ to $\hat{\Delta}$ implies that the image of $\sigma(f, \varphi)$ cannot intersect $\hat{\Delta}$. But this is precisely the condition for $\Phi_{f,\varphi}$ to be 1-1, and hence if $d \geq 2m + 1$, injective maps are residual in $\mathcal{E}^r$ and hence residual in $\mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$.

This completes the proof of Takens Theorem. We make one final comment: In the proofs of the transversality of ev_ρ and ev_σ below, we only make use of perturbations in φ. Thus for a fixed $f \in \mathcal{D}^r$, ev_ρ and ev_σ will be transversal to L and $\hat{\Delta}$, respectively, when considered as functions of φ only. This immediately gives a proof of a version of the "unstated" Takens Theorem ([Huke, 1993]):

Theorem 4.7. *If $f \in \mathcal{D}^r$, then there is an open and dense set of observation functions $\varphi \in \mathcal{C}^r(M, \mathbb{R})$ for which $\Phi_{f,\varphi}$ is an embedding.*

Note that this is slightly weaker than Theorem 2.2, since the definition of $\mathcal{D}^r$ places restrictions on periodic orbits of all periods up to $2d$, rather than just d, which is the case in Theorem 2.2.

4.2. Embedding the Short Periodic Orbits

4.2.1. Isolating the Periodic Points. Although Proposition 4.1 is a simple corollary of the Kupka-Smale Theorem, we give a detailed proof. This is because very similar ideas arise in Section 5 in the proof of the Forced Takens Theorem in a context where we cannot deduce the required result from the Kupka-Smale Theorem. It thus seems helpful to introduce the techniques that we shall use there, within the much simpler setting of the unforced case. The proof is actually our first application of the Parametric Transversality Theorem.

Define

$$\mathcal{B}_q = \{ f \in \mathcal{D}^r(M)\colon \text{ all periodic orbits of period } k < q \text{ are isolated and hyperbolic}\},$$

with $\mathcal{B}_1 = \mathcal{D}^r(M)$ by default.

Lemma 4.8. *$\mathcal{B}_q$ is open and dense in $\mathcal{D}^r(M)$ for all $q \geq 1$, $r \geq 2$.*

Proof. By induction. Suppose the lemma holds for some $q \geq 1$. Then $\mathcal{B}_q$ is a Banach manifold and we may apply the Parametric Transversality Theorem to $\rho\colon \mathcal{B}_q \to \mathcal{C}^r(M, M\times M)$ given by $\rho(f) = (f^q, id)$. The evaluation map is $ev_\rho(f, x) = (f^q(x), x)$. By Corollary C.2, ev_ρ is $\mathcal{C}^1$. The periodic points of f of period q correspond to points of intersection of the image of $\rho(f)$ with the diagonal $\Delta = \{(x, x)\colon\ x \in M\} \subset M \times M$. This is a closed submanifold of $M \times M$ of codimension m, with $T_{x,x}\Delta = \{(v, v)\colon\ v \in T_x M\}$. To show that ev_ρ is transversal to Δ at (f, x); it is sufficient to show that if $ev_\rho(f, x) \in \Delta$ then the image of $T_{f,x}(ev_\rho)$ contains the space $\{(u, 0)\colon\ u \in T_x M\}$. By Corollary C.2 we have

$$T_{f,x}(ev_\rho)(\eta, 0_x) = \left(\sum_{i=1}^{q} T_{x_i} f^{q-i}(\eta(x_{i-1})), 0_x\right)$$

where $T_{x_q}(f^0) = Id$. First suppose that x has least period q. Then the points $x_i = f^i(x)$, $i = 0, \ldots, q-1$ are disjoint, and by Corollary C.12, given any $u \in T_x M$, we can find a $\eta \in T_f \mathcal{B}_q$ such that $\eta(x_{q-1}) = u \in T_{x_q} M = T_x M$ and $\eta(x_i) = 0_{x_{i+1}}$ for $i = 0, \ldots, q-2$. For such a η, we have $T_{f,x}(ev_\rho)(\eta, 0) = (u, 0)$, and hence $T_{f,x}(ev_\rho)(T_{f,x}(\mathcal{B}_q \times M))$ contains the space $\{(u, 0)\colon\ u \in T_x M\}$ as required. Now suppose that x has least period $p < q$. Then

$$T_{f,x}(ev_\rho)(\eta, 0_x) = \left(\sum_{i=1}^{p} T^{(k)}(T_{x_i} f^{p-i}(\eta(x_{i-1}))), 0_x\right),$$

where

$$T^{(k)} = \sum_{j=1}^{k} T_x f^{p(j-1)},$$

where $q = kp$. Since x is periodic of period k, none of the eigenvalues of $T_x f^p$ have unit modulus (by the inductive hypothesis). The eigenvalues of $T^{(k)}$ are of the form $1+\lambda+\cdots+\lambda^{k-1}$, for λ an eigenvalue of $T_x f^p$. Since $(1+\lambda+\cdots+\lambda^{k-1})(1-\lambda) = (1-\lambda^k)$, we see that all of the eigenvalues of $T^{(k)}$ are nonzero, and hence $T^{(k)}$ is invertible. Now we proceed as before: The points $x_i = f^i(x)$, $i = 0, \ldots, p-1$ are distinct, so given any $u \in T_x M$, we can find a $\eta \in T_f \mathcal{B}_q$ such that $\eta(x_{p-1}) = (T^{(k)})^{-1} u \in T_{x_p} M = T_x M$ and $\eta(x_i) = 0_{x_{i+1}}$ for $i = 0, \ldots, p-2$. For such a η, we have $T_{f,x}(ev_\rho)(\eta, 0) = (u, 0)$, as required.

We have thus shown that ev_ρ is transversal to the diagonal Δ in $M \times M$. Since Δ is a closed submanifold of $M \times M$ of codimension m, the Parametric Transversality Theorem implies that the set of $f \in \mathcal{D}^r(M)$ for which (f^q, id) is transversal to Δ is open and dense for $r > 0$. But if (f^q, id) is transversal to Δ, then $(f^q, id)^{-1}(\Delta)$ is a submanifold of M of codimension m, and hence of dimension 0. It is thus a finite set of isolated points (recall M is compact). But $(f^q, id)^{-1}(\Delta)$ is nothing more than the set of periodic points of f of period q. To complete the inductive step we need to show that we can further restrict the set of f to make the periodic orbits of period q hyperbolic. Suppose that $f_0 \in \mathcal{B}_q$ such that $((f_0)^q, id)$ is transversal to Δ. Then by the Transversal Isotopy Theorem (Appendix A) there is an open neighbourhood $\mathcal{U}$ of f_0 such that every f in $\mathcal{U}$ has the same number of periodic points of period q as f_0. We can thus perturb f_0 by an arbitrarily small amount in the neighbourhood of each such periodic orbit (using Corollary C.18 if we want to be completely rigorous) to give an f

such that these periodic orbits are hyperbolic. Thus $\mathcal{B}_{q+1}$ is dense in $\mathcal{B}_q$. Conversely, if for some f_0 all the periodic orbits of period q are already hyperbolic, then they will be hyperbolic for all sufficiently close f and hence $\mathcal{B}_{q+1}$ is open in $\mathcal{B}_q$. We have thus shown that $\mathcal{B}_{q+1}$ is open and dense in $\mathcal{B}_q$, and since $\mathcal{B}_q$ is open and dense in $\mathcal{D}^r(M)$, $\mathcal{B}_{q+1}$ is also open and dense in $\mathcal{D}^r(M)$, thus completing the inductive step. □

Proof of Proposition 4.1. Observe that we can trivially modify the last step in the proof of Lemma 4.8 to require the eigenvalues of any periodic orbit of period q to be distinct. Taking the intersection of all such sets for $q < 2d$, we obtain precisely $\mathcal{D}^r$, and since the intersection of a finite number of open and dense sets is open and dense, this completes the proof. □

4.2.2. Injectivity on the Periodic Points.

Proof of Proposition 4.2. The set $\mathcal{P}_f$ consists of a finite number of points; write these as $x_0, \ldots, x_{k-1}$. Let ρ: $\mathcal{C}^r(M, \mathbb{R}) \to \mathbb{R}^k$ be given by $\rho(\varphi) = (\varphi(x_0), \ldots, \varphi(x_{k-1}))$. Let Δ_k be the open and dense subset of $\mathbb{R}^k$ given by $\Delta_k = \{(y_0, \ldots, y_{k-1}) \in \mathbb{R}^k\colon\ y_i \neq y_j$ if $i \neq j\}$. Then $\rho^{-1}(\Delta_k)$ is precisely the set of φ such that φ is 1-1 on $\mathcal{P}_f$. By Corollary B.3 the map ρ is continuous and hence $\rho^{-1}(\Delta_k)$ is open in $\mathcal{C}^r(M, \mathbb{R})$. By Corollary C.13 ρ is a submersion, and by a version of the implicit function theorem (e.g., see Abraham and Robbin [1967] or Lang [1972]) any submersion is a local fibration, i.e., there are local coordinates $\varphi = (\varphi_1, \varphi_2)$ on $\mathcal{C}^r(M, \mathbb{R})$ such that ρ is locally a projection onto the first factor. Therefore $\rho^{-1}(\Delta_k)$ is also dense, as required. □

4.2.3. Immersing the Periodic Points.

Proof of Proposition 4.3. Let $V_i^* = \mathcal{L}(T_x M, T_{\varphi(x_i)}\mathbb{R})$ and define $a_i \in V_i^*$ by $a_i = T_{x_i}\varphi \circ T_x f^i$ for any i. Then the i^{th} component of $T_x \Phi_{f,\varphi}(v)$ is $a_i(v)$, and to show that $T_x \Phi_{f,\varphi}$ has rank m it is sufficient to show that $a_0 \wedge \cdots \wedge a_{m-1} \neq 0$.

Suppose that x has minimal period q. Since $x_0, \ldots, x_{q-1}$ are distinct, the maps $T_{x_0}\varphi, \ldots, T_{x_{q-1}}\varphi$ can be adjusted independently, or more precisely by Corollary C.18 the map $\varphi \mapsto (T_{x_0}\varphi, \ldots, T_{x_{q-1}}\varphi)$ is a submersion. But $T_x f^i$ is a linear isomorphism for each i, and therefore the map ρ: $\mathcal{C}^r(M, \mathbb{R}) \to V_0^* \times \cdots \times V_{q-1}^*$ given by $\rho(\varphi) = (T_{x_0}\varphi, \ldots, T_{x_{q-1}}\varphi \circ T_x f^{q-1}) = (a_0, \ldots, a_{q-1})$ is also a submersion. The fact that $a_0 \wedge \cdots \wedge a_{m-1} \neq 0$ for an open and dense set in $\mathcal{C}^r(M, \mathbb{R})$ then follows immediately from the following.

Lemma 4.9 (Huke, [1993]). *Let V be a vector space with* $\dim V = m$. *Let A: $V \to V$ be an invertible linear map with distinct eigenvalues. Let $m = qr + s$, with $0 < s \leq q$, and let $V_0, \ldots, V_{q-1}$ be any collection of one-dimensional linear spaces. Define $V_i^* = \mathcal{L}(V, V_i)$. For any $a = (a_0, \ldots, a_{q-1}) \in V_0^* \times \cdots \times V_{q-1}^*$ let $\omega_a \in V_0^* \wedge \cdots \wedge V_{m-1}^*$ be given by $\omega_a = a_0 \wedge \cdots \wedge a_{q-1} \wedge a_0 \circ A \wedge \cdots \wedge a_{q-1} \circ A \wedge \cdots \wedge a_0 \circ A^{r-1} \wedge$*

$\cdots \wedge a_{q-1} \circ A^{r-1} \wedge a_0 \circ A^r \wedge \cdots \wedge a_{s-1} \circ A^r$. *Then* $\omega_a \neq 0$ *for an open dense set of* $(a_0, \ldots, a_{q-1}) \in V_0^* \times \cdots \times V_{q-1}^*$.

Proof. Let $e_0, \ldots, e_{m-1}$ be a basis of V. Since $V_0^* \wedge \cdots \wedge V_{m-1}^*$ is one-dimensional, $\omega_a \neq 0$ if an only if $\omega_a(e_0, \ldots, e_{m-1}) \neq 0$. Now, $\omega_a(e_0, \ldots, e_{m-1}) = \det B$, where B is an $m \times m$ matrix such that $B_{ij} = a_k \circ A^p(e_j)$ where $i = qp + k$, with $0 \leq k < q$. Thus B_{ij} is a linear combination of $B_{k,0}, \ldots, B_{k,m-1}$, and hence $\det B$ is a polynomial in the qm variables $B_{0,0}, \ldots, B_{0,m-1}, \ldots, B_{q-1,0}, \ldots, \mathcal{B}_{q-1,m-1}$. Since the map $a_k \mapsto B_{k,0}, \ldots, B_{k,m-1}$ is an isomorphism of V_k^* and $\mathbb{R}^m$, we need to show that this polynomial is nonzero for an open dense set of $(B_{0,0}, \ldots, B_{q-1,m-1})$ in $\mathbb{R}^{qm}$.

Recall that the set of roots $\beta^{-1}(0)$ of any polynomial β which does not vanish identically is a closed nowhere dense set. This is because $\beta^{-1}(0)$ is trivially closed since α is continuous and it must have empty interior: If β were identically zero on some open neighbourhood, then all its derivatives there would vanish; hence its Taylor expansion would vanish identically. But this Taylor expansion is just β itself, and hence β would vanish everywhere.

It thus remains to construct a single $(a_0, \ldots, a_{q-1})$ such that $\omega_a(e_0, \ldots, e_{m-1}) \neq 0$. This is done as follows: Choose any a_0 such that $a_0(e_j) \neq 0$ for all $j = 0, \ldots, m-1$. For $i \leq s$ let $a_i = a_0 \circ A^{i(r+1)}$ and for $s < i \leq q-1$ let $a_i = a_0 \circ A^{s+ir}$. The a_i for $i \geq q$ are then of course given by $a_i + k_q = a_i \circ A$ for any i and k. Note that if $s < q$ then $a_{q-1} \circ A^{r-1} = a_0 \circ A^{m-1}$ and if $s = q$ then $a_{s-1} \circ A^r = a_0 \circ A^{m-1}$. Thus $a_0, \ldots, a_{m-1}$ are a permutation of $a_0, \ldots, a_0 \circ A^{m-1}$ and hence $\omega_a = \pm a_0 \wedge a_0 \circ A \wedge \cdots \wedge a_0 \circ A^{m-1}$. But by Lemma 4.10 below, $a_0, a_0 \circ A, \ldots, a_0 \circ A^{m-1}$ are linearly independent if $a_0(e_j) \neq 0$ for all $j = 1, \ldots, m$. Thus $\omega_a(e_0, \ldots, e_{m-1}) \neq 0$, as required. □

Lemma 4.10. *Let* V *be a vector space with* $\dim V = m$. *Let* $A\colon V \to V$ *be an invertible linear map with distinct eigenvalues,* $a\colon V \to \mathbb{R}$ *be any linear map such that* $a(v) \neq 0$ *for all eigenvectors* v *of* A. *Then the set* $a, a \circ A, \ldots, a \circ A^{m-1}$ *are linearly independent.*

Proof. Suppose not. Then there exists $\alpha_0, \ldots, \alpha_{m-1}$ with at least one $\alpha_i \neq 0$, such that

$$\sum_{j=0}^{m-1} \alpha_j a \circ A^j = 0.$$

Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of A. Since the λ_i are distinct we can find a basis of V consisting of eigenvectors $v_1, \ldots, v_m$ of A (if some of the λ_i are complex, we simply work with the complexification of V). Then for all $i = 1, \ldots, m$ we have

$$\sum_{j=0}^{m-1} \alpha_j a \circ A^j(v_i) = 0.$$

But $A^j(v_i) = (\lambda_i)^j v_i$ and so for any $i = 1, \ldots, m$,

$$\begin{aligned} \sum_{j=0}^{m-1} \alpha_j a \circ A^j(v_i) &= a(v_i) \sum_{j=0}^{m-1} \alpha_j (\lambda_i)^j \\ &= a(v_i) p(\lambda_i), \end{aligned}$$

where $p(\lambda)$ is the polynomial $p(\lambda) = \alpha_0 + \alpha_1\lambda + \cdots \alpha_{m-1}(\lambda)^{m-1}$. But $a(v_i) \neq 0$ since v_i is an eigenvector v of A and hence $p(\lambda_i) = 0$ for all $\lambda_1, \ldots, \lambda_m$. The λ_i are distinct and thus $p(\lambda)$ has m distinct roots. This contradicts the fact that it is a nontrivial polynomial of degree at most $m-1$, and hence can have at most $m-1$ roots. □

4.2.4. Proof of Lemma 4.4. The density of $\mathcal{E}^r$ follows immediately from that of $\mathcal{D}^r$ and $\mathcal{A}_f$. Thus given any $(f_0, \varphi_0) \in \mathcal{D}^r(M) \times \mathcal{C}^r(M, \mathbb{R})$, first perturb f_0 by an arbitrarily small amount to give an $f \in \mathcal{D}^r$, and then make an equally small perturbation to give a $\varphi \in \mathcal{A}_f$.

To show that $\mathcal{E}^r$ is also open, note that if $f_0 \in \mathcal{D}^r$ then every f in the neighbourhood f_0 has the same number of periodic orbits of period less than d as f_0, and the positions of the orbits depend continuously on f. Thus if φ_0 is 1-1 on $\mathcal{P}_{f_0}$; then φ will be 1-1 on $\mathcal{P}_f$ for all (f, φ) sufficiently close to (f_0, φ_0). Furthermore, $T_x\Phi_{f,\varphi}$ depends continuously on x, f, and φ, and hence if Φ_{f_0,φ_0} is immersive on $\mathcal{P}_{f_0}$ then $\Phi_{f,\varphi}$ is immersive on $\mathcal{P}_f$ for all (f, φ) in a sufficiently small neighbourhood of (f_0, φ_0). □

4.3. Transversality of ev_ρ

Proof of Proposition 4.5. If $ev_\rho(f, \varphi, v) \in L$ with $(f, \varphi) \in \mathcal{E}^r$, $v \in \tilde{T}_x M$, then x cannot lie on a periodic orbit of period $< d$. This is because if $(f, \varphi) \in \mathcal{E}^r$, then by definition if x is a periodic orbit of period $< d$, then $T_x\Phi_{f,\varphi}$ has rank m and hence $T_x\Phi_{f,\varphi}(v) \neq 0$ for all $v \neq 0$. Thus ev_ρ is trivially transversal to L at such periodic x.

By Corollary C.8,

$$T_{f,\varphi,v}(ev_\rho)(0_f, \xi, 0_v) = \begin{pmatrix} \omega(T_x\xi(v)) \\ \omega(T_{x_1}\xi(v_1)) \\ \vdots \\ \omega(T_{x_{d-1}}\xi(v_{d-1})) \end{pmatrix},$$

where $x_i = f^i(x)$ and $v_i = T_x f^i(v)$. Denote $v' = T_x\Phi_{f,\varphi}v$. If x is not a periodic orbit of period $< d$, then the points x_i, $i = 0, \ldots, d-1$, are distinct. Also, since $v \neq 0$, and f is a diffeomorphism, we have $v_i \neq 0$ for all $i = 0, \ldots, d-1$. Hence by Corollary C.16, given any $(u_1, \ldots, u_{d-1}) = u \in T_{v'}(T\mathbb{R}^d)$ there exists a $\xi \in T_\varphi(\mathcal{C}^r(M, \mathbb{R}))$ such that $\omega(T_{x_i}\xi(v_i)) = u_i$ for $i = 0, \ldots, d-1$, or in other words such that $T_{f,\varphi,v}(ev_\rho)(0_f, \xi, 0_v) = u$. Thus $T_{f,\varphi,v}(ev_\rho)$ is surjective, and hence it is transversal to any submanifold of $T\mathbb{R}^d$; in particular it is transversal to L as required. □

4.4. Transversality of ev_σ

Proof of Proposition 4.6. Suppose that $ev_\sigma(f, \varphi, x, x') \in \hat{\Delta}$ with $(f, \varphi) \in \mathcal{E}^r$. Then $\Phi_{f,\varphi}(x) = \Phi_{f,\varphi}(x')$, and hence $\varphi(x) = \varphi(x')$, but $x \neq x'$. Thus, at least one of x or x' is not a periodic point of period $< 2d$, since if $(f, \varphi) \in \mathcal{E}^r$, φ is injective on the set of such periodic points. Thus, without loss of generality, we may assume that the points $x_i = f^i(x)$, $i = 0, \ldots, d-1$, are distinct. Let $z = \Phi_{f,\varphi}(x) = \Phi_{f,\varphi}(x')$, then $T_{z,z}\hat{\Delta} = \{(u, u)\colon u \in T_y\mathbb{R}^d\}$. We need to show that the image of $T_{f,\varphi,x,x'}(ev_\sigma)$ contains a complement to this space. Note that $\xi(x_i), \xi(x_i') \in T_{z_i}\mathbb{R}$, which is a one-dimensional space. Let e_i be a basis vector for $T_{y_i}\mathbb{R}$, and let $e^{(i)}$ be the vector $(0, \ldots, e_i, \ldots, 0)^\dagger \in T_y\mathbb{R}^d$. Then $e^{(0)}, \ldots, e^{(d-1)}$ forms a basis for $T_z\mathbb{R}^d$. For ease of notation write $\Xi = (f, \varphi, x, x')$. Then by Corollary C.3,

$$T_\Xi(ev_\sigma)(0_f, \xi, 0_x, 0_{x'}) = \left(\begin{pmatrix} \xi(x) \\ \xi(x_1) \\ \vdots \\ \xi(x_{d-1}) \end{pmatrix}, \begin{pmatrix} \xi(x') \\ \xi(x_1') \\ \vdots \\ \xi(x_{d-1}') \end{pmatrix} \right).$$

First suppose that $\{x, x_1, \ldots, x_{d-1}\} \cap \{x', x_1', \ldots, x_{d-1}'\} = \varnothing$. Then using Corollary C.12, for each $i = 0, \ldots, d-1$ we can find a $\xi_i \in T_\varphi(\mathcal{C}^r(M, \mathbb{R}))$ such that $\xi_i(x_i) = e_i$, $\xi_i(x_j) = 0$ for $j \neq i$, and $\xi_i(x_i) = 0$ for all $j = 0, \ldots, d-1$. For such a ξ_i we have $T_\Xi(ev_\sigma)(0_f, \xi_i, 0_x, 0_{x'}) = (e^{(i)}, 0_z)$, and hence the image of $T_\Xi(ev_\sigma)$ contains the space $T_z\mathbb{R}^d \times \{0_y\}$, which is clearly a complement of $T_{z,z}\hat{\Delta}$, as required.

It remains to deal with the case $x' = f^j(x)$ for some $-(d-1) \leq j \leq d-1$, $j \neq 0$. Since x' cannot then be a periodic orbit of period $< 2d$ we can assume without loss of generality (interchanging x and x' if necessary) that $x' = f^j(x)$ for some $0 < j \leq d-1$. Note that if $f^i(x') = x$ for some $i > 0$, then $i + j \geq 2d$, since x is not periodic with period less than $2d$. Hence $\{x, x_1, \ldots, x_{j-1}\} \cup \{x', x_1' \ldots, x_{d-1}'\}$ are disjoint. Hence for $0 \leq i < j$ we can find as above a $\xi_i \in T_\varphi(\mathcal{C}^r(M, \mathbb{R}))$ such that $T_\Xi(ev_\sigma)(0_f, \xi_i, 0_x, 0_{x'}) = (e^{(i)}, 0_z)$. Now proceed by induction. Our inductive hypothesis on k is that for all $0 \leq i < k$ there exists a $\xi_i \in T_\varphi(\mathcal{C}^r(M, \mathbb{R}))$ and a $u_i \in T_z\mathbb{R}^d$ such that $T_\Xi'(ev_\sigma)(0_f, \xi_i, 0_x, 0_{x'}) = (e_i + u_i, u_i)$. By the above this holds for $k = j - 1$ (with all the $u_i = 0_z$).

The inductive step then proceeds as follows: Since $f^{i-j}(x') = f^i(x)$, the points $\{x, x_1, \ldots, x_{d-1}\} \cup \{x_{d-j}', \ldots, x_{d-1}'\}$ are disjoint, so we can find a $\zeta_k \in T_\varphi(\mathcal{C}^r(M, \mathbb{R}))$ such that $\zeta_k(x_k) = e_k$, $\zeta_k(x_i) = 0$ for $i \neq k$, and $\zeta_k(x_i') = 0$ for $i = d-j, \ldots, d-1$. The remaining values $\zeta_k(x_i')$ for $i = 0, \ldots, d-j-1$ are determined by $x_i' = f^i(x') = f^{i+J}(x) = x_{i+j}$, so that $\zeta_k(x_{k-j}') = \zeta_k(x_k) = e_k$ and $\zeta_k(x_i') = \zeta_k(x_{i+j}) = 0$ for $i \neq k-j$. Putting all this together we get $T_\Xi(ev_\sigma)(0_f, \zeta_k, 0_x, 0_{x'}) = (e_k, e_{k-j})$. By the inductive hypothesis there exists a ξ_{k-j} such that $T_\Xi(ev_\sigma)(0_f, \xi_{k-j}, 0_x, 0_{x'}) = (e_{k-j} + u_{k-j}, u_{k-j})$. Thus $T_\Xi(ev_\sigma)(0_f, (\zeta_k + \xi_{k-j}), 0_x, 0_{x'}) = (e_k + e_{k-j} + u_{k-j}, e_{k-j} + u_{k-j})$. This completes the inductive step with $u_k = e_{k-j} + u_{k-j}$.

We have thus shown that the image of $T_\Xi(ev_\sigma)$ contains a vector of the form $(e_i + u_i, u_i)$ for all $i = 0, \ldots, d-1$. But $(u_i, u_i) \in T_{z,z}\hat{\Delta}$, and hence $(e^{(i)}, 0_z) \in \mathrm{Im}(T_\Xi(ev_\sigma)) + T_{z,z}\hat{\Delta}$. Thus $\mathrm{Im}(T_\Xi(ev_\sigma)) + T_{z,z}\hat{\Delta}$ contains $T_z\mathbb{R}^d \times \{0_y\}$, which is a complement of $T_{z,z}\hat{\Delta}$, as required. □

5. Proof of Takens Theorem for Skew Products

This follows very similar lines to the proof of the standard Takens Theorem given in the previous section. First, observe that since the map $(f, \varphi) \mapsto \Phi_{f,g,\varphi}$ is continuous, the set of (f, φ) such that $\Phi_{f,g,\varphi}$ is an embedding is open. Secondly, since $\mathcal{C}^r(M, N)$ is dense in $\mathcal{C}^{r'}(M, N)$ for all $r > r'$ we only need to prove the theorem for all sufficiently large r. It turns out we shall need $r \geq 2d$.

As in the standard Takens Theorem, the main problem is dealing with points such that $x_i = x_j$ with $i \neq j$. Of course x_i is now defined by $x_i = f^{(i)}(x, y)$, and hence there are far more such points than just the periodic points of (f, g), as was the case in Section 4. This is the fundamental issue that makes the proof of the forced version of the theorem so much more delicate than that of the standard version.

Our approach here is to ensure that these problem points occur on families of submanifolds W_I of $M \times N$ and then to deal with each of these separately. Each W_I is characterized by the set of pairs (i, j) for which $x_i = x_j$. The key point is that the more such pairs (i, j) there are, the smaller the dimension of the corresponding W_I, and hence the fewer independent components of $\Phi_{f,g,\varphi}$ are required to embed it. It thus turns out that as long as $\dim M \geq 1$, the loss of independence due to each relation $x_i = x_j$ is compensated for by the drop in the dimension of W_I.

We shall construct the W_I (and ensure that they are submanifolds) by using the Parametric Transversality Theorem to get a residual set of f for which the map $\tau(f) = (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(d-1)})\colon M \times N \to M^d$ is transversal to an appropriate submanifold of M^d. A straightforward argument (Lemma 5.12 below, which itself is a simple consequence of Corollary C.5) shows that ev_τ is submersive except at periodic points of period less than d. It turns out that W_I need not be a manifold at such points. This is to be expected, since at periodic points, the condition $x_i = x_j$ immediately implies $x_i + k = x_j + k$ for any k, and hence relations of the form $x_i = x_j$ are not independent. We thus first have to exclude the periodic points of period $q < d$ from W_I and deal with them separately, just as in the standard Takens Theorem.

Finally, we remark that since the W_I are constructed by a transversality argument, we could avoid explicit construction of these manifolds by combining this argument with the proof of the transversality of $\tilde{T}\Phi_{f,g,\varphi}$ and $\Phi_{f,g,\varphi} \times \Phi_{f,g,\varphi}$. Whilst this leads to a slightly shorter proof, we prefer not to use this approach here since we believe that the construction of the W_I (and their generalizations) gives much more geometric insight into the principles behind the proof. It turns out, however, that this unified approach is necessary in the proof of Theorem 3.2 in the next section, and hence the interested reader is referred to that section to see how such an integrated argument works.

5.1. Main Proof

5.1.1. Periodic Orbits. Our first task, as in Section 4, is to show that for a dense open set of f, the periodic orbits of any given period are isolated, and each one has distinct eigenvalues. In proving injectivity at pairs of points (x, y), (x', y') such that $(x', y') = (f, g)^k(x, y)$, it will turn out to be necessary to require this for all periods less than $2d$. We thus define

$$\begin{aligned}\mathcal{D}^r = \{ f \in \mathcal{D}^r(M \times N, M)\colon & \text{ all periodic points of } (f, g) \text{ of period } q < 2d \\ & \text{are isolated and hyperbolic, have distinct } x \\ & \text{coordinates and each has distinct eigenvalues}\}.\end{aligned}$$

In Section 5.2 we prove the following:

Proposition 5.1. *$\mathcal{D}^r$ is open and dense in $\mathcal{D}^r(M \times N, M)$ for $r \geq 2$. Furthermore, for* *any $f \in \mathcal{D}^r$ there is an open neighbourhood in $\mathcal{D}^r$ (and hence $\mathcal{D}^r(M \times N, M)$) such that all maps in this neighbourhood have the same number of periodic orbits of each period less than $2d$.*

For any $f \in \mathcal{D}^r$ let $\mathcal{P}^{(r)} f, g$ be the set of periodic points of (f, g) of period less than r, for $1 \leq r \leq 2d$, and for convenience denote $\mathcal{P}_{f,g} = \mathcal{P}^{(d)}_{f,g}$. We shall first show that $\Phi_{f,g,\varphi}$ is 1-1 on $\mathcal{P}^{(2d)}_{f,g}$ for an open dense set in $\mathcal{D}^r \times \mathcal{C}^r(M, \mathbb{R})$.

Since all points in $\mathcal{P}^{(2d)}_{f,g}$ have different x coordinates, the projection $\pi_1\colon M \times N \to M$ is 1-1 on $\mathcal{P}^{(2d)}_{f,g}$. By a similar argument to the proof of Proposition 4.2 (in Section 4.2.2), the set of $\varphi \in \mathcal{C}^r(M, \mathbb{R})$ that take distinct values on $\pi_1(\mathcal{P}^{(2d)}_{f,g})$ is open and dense in $\mathcal{C}^r(M, \mathbb{R})$. Call this set $\mathcal{A}_{f,g}$. For any $\varphi \in \mathcal{A}_{f,g}$, the function $\varphi \circ \pi_1$ is 1-1 on $\mathcal{P}^{(2d)}_{f,g}$. But $\varphi \circ \pi_1$ is precisely the first component of $\Phi_{f,g,\varphi}$, and hence $\Phi_{f,g,\varphi}$ is also 1-1 for all such φ. Then, as in the proof of Lemma 4.4, it is easily seen that the set of (f, φ) such that $\Phi_{f,g,\varphi}$ is also 1-1 on $\mathcal{P}^{(2d)}_{f,g}$ is open and dense in $\mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$.

Now let us turn to the immersivity of $\Phi_{f,g,\varphi}$ on the periodic orbits. It turns out we only need consider periods up to d in this case. Fix some $(f_0, \varphi_0) \in \mathcal{D}^r \times \mathcal{C}^r(M, \mathbb{R})$. We will construct an (f, φ) arbitrarily close to (f_0, φ_0) for which $\Phi_{f,g,\varphi}$ is immersive on $\mathcal{P}_{f,g}$. It will turn out to be convenient for f to have the same periodic orbits as f_0; thus define

$$\mathcal{D}^r(\mathcal{P}_{f_0,g}) = \{f \in \mathcal{D}^r\colon\ f(x, y) = f_0(x, y) \text{ for all } (x, y) \in \mathcal{P}_{f_0,g}\}.$$

By Corollary C.13 this is a submanifold of $\mathcal{D}^r$. Similarly let

$$\mathcal{C}^r(M, \mathbb{R}; \mathcal{P}_{f_0,g}) = \{\varphi \in M, \mathbb{R}\colon\ \varphi(x) = \varphi_0(x) \text{ for all } x \in \pi_1(\mathcal{P}_{f,g})\}.$$

A rather long and technical argument in Section 5.2 below gives the following.

Proposition 5.2. *Given $(f_0, \varphi_0) \in \mathcal{D}^r \times \mathcal{C}^r(M, \mathbb{R})$ and any $(x, y) \in \mathcal{P}_{f_0,g}$, there is an open dense set in $\mathcal{D}^r(\mathcal{P}_{f_0,g}) \times \mathcal{C}^r(M, \mathbb{R}; \mathcal{P}_{f_0,g})$ such that for all (f, φ) in this set $T_{x,y}\Phi_{f,g,\varphi}$ is 1-1.*

Observe that by Proposition 5.1, for any $f \in \mathcal{D}^r(\mathcal{P}_{f_0,g})$ sufficiently close to f_0 we have $\mathcal{P}_{f,g} = \mathcal{P}_{f_0,g}$. Hence the set of (f, φ) such that $\Phi_{f,g,\varphi}$ is immersive on $\mathcal{P}_{f,g}$ is dense. This set is also open, by the same argument as in Section 4.2.4. Thus in a sufficiently small neighbourhood of f_0 in $\mathcal{D}^r$, the positions of the points in $\mathcal{P}_{f,g}$ depend continuously on f. Since $T_{x,y}\Phi_{f,g,\varphi}$ depends continuously on x, y, f, and φ, if Φ_{f_0,g,φ_0} is immersive on $\mathcal{P}_{f_0,g}$ then $\Phi_{f,g,\varphi}$ is immersive on $\mathcal{P}_{f,g}$ for all (f, φ) in a sufficiently small neighbourhood of (f_0, φ_0). We thus have

Corollary 5.3. *The set of $(f, \varphi) \in \mathcal{D}^r \times \mathcal{C}^r(M, \mathbb{R})$ such that $\Phi_{f,g,\varphi}$ is immersive for all $(x, y) \in \mathcal{P}_{f,g}$ is open and dense in $\mathcal{D}^r \times \mathcal{C}^r(M, \mathbb{R})$.*

5.1.2. The Sets $\tilde{W}_I$. Having dealt with the periodic orbits, we next show how to construct the W_I. Let $I = \{I_1, I_2, \ldots, I_\alpha\}$ be a partition of $\{0, \ldots, d-1\}$, and define the associated equivalence relation $\sim_I$ on $\{0,\ldots,d-1\}$ by $i \sim_I i'$ if and only if i, i' are in the same element of the partition. We then define the set W_I by

$$W_I = \{(x, y) \in M \times N\colon\ x_i = x_{i'} \text{ if and only if } i \sim_I i'\}.$$

Recall that for any $f \in \mathcal{D}^r$, the set $\mathcal{P}_{f,g}$ of periodic points of period less than d is finite, and hence $(M \times N)\backslash\mathcal{P}_{f,g}$ is an open submanifold of $M \times N$. Define

$$\tilde{W}_I = W_{I\cap}(M \times N)\backslash\mathcal{P}_{f,g}.$$

Then,

Proposition 5.4. *For a residual set of $f \in \mathcal{D}^r$, if $\tilde{W}_I$ is nonempty then it is a codimension $(d-\alpha)m$ submanifold of $(M \times N)\backslash\mathcal{P}_{f,g}$.*

The proof is given in Section 5.3. Given any partition I as above, let J_I be a set containing precisely one element from each I_k for $k = 1,\ldots,\alpha$. There will typically be many ways to choose such a J_I, but we arbitrarily select just one. Clearly J_I has α elements. Write these as $J_I = \{j_1, j_2, \ldots, j_\alpha\}$ with $j_1 < j_2 < \cdots < j_\alpha$. Note that if $(x, y) \in \tilde{W}_I$, then the points $x_{j_1}, x_{j_2}, \ldots, x_{j_\alpha}$ are distinct and any other x_j for $j = 0,\ldots,d-1$ is equal to one of these. In other words, the set $\{x_0,\ldots,x_{d-1}\}$ contains precisely α distinct points. Now, $2\dim \tilde{W}_I + 1 = 2(m + n - (d-\alpha)m) + 1 \le d - 2(d-\alpha)m = (d-\alpha)(1-2m) + \alpha \le \alpha$. Thus it seems plausible that generically $\Phi_{f,g,\varphi}$ should embed each $\tilde{W}_I$. It turns out however that the $\tilde{W}_I$ are not quite the right object to consider, since to show immersivity of $\Phi_{f,g,\varphi}$ we need to work in the unit tangent bundle $\tilde{T}(M\times N)$ of $M\times N$, and to show injectivity we need to work with $(M\times N)\times(M\times N)$. We thus need to define appropriate generalizations of $\tilde{W}_I$ that give decompositions of $\tilde{T}(M \times N)$ and $(M \times N) \times (M \times N)$, respectively.

5.1.3. Immersivity.

We first consider $\tilde{T}(M\times N)$. Let $I' = \{i_1, i_2, \ldots, i_\beta\}$ be any subset of J_I (possibly empty). Let $\tilde{T}_{\tilde{W}_I}((M \times N)\backslash\mathcal{P}_{f,g})$ be the restriction of the unit tangent bundle of $(M \times N)\backslash\mathcal{P}_{f,g}$ to $\tilde{W}_I$. Define the set

$$\tilde{T}_{I,I'} = \{v \in \tilde{T}_{\tilde{W}_I}((M \times N)\backslash\mathcal{P}_{f,g})\colon \text{ if } i \in J_I, \text{ then } \tilde{T} f^{(i)} v = 0_{x_i} \text{ if and only if } i \in I'\}.$$

Thus I' specifies those points in $x_{j_1}, x_{j_2}, \ldots, x_{j_\alpha}$ where $\tilde{T} f^{(i)}$ vanishes. We do not care what $\tilde{T} f^{(i)}$ does for $i \notin J_I$, since we cannot perturb φ independently at the corresponding x_i. The union of the $\tilde{T}_{I,I'}$ over all I' gives $\tilde{T}_{\tilde{W}_I}((M \times N)\backslash\mathcal{P}_{f,g})$ (recall that we include the case $I' = \varnothing$) and hence the union of the $\tilde{T}_{I,I'}$ over all I and I' is $\tilde{T}((M \times N)\backslash\mathcal{P}_{f,g})$. Let $\gamma = \alpha - \beta$. In Section 5.4 below, we prove the following.

Proposition 5.5. *For a residual set of $f \in \mathcal{D}^r$, if $\tilde{T}_{I,I'}$ is nonempty then it is a codimension $(d-\gamma)m$ submanifold of $\tilde{T}((M \times N)\backslash\mathcal{P}_{f,g})$.*

By taking the intersection over all I and I' we get a residual set of f for which $\tilde{T}_{I,I'}$ is a submanifold for all I and I'. Fix an f in this set. Note that $J_I\backslash I'$ has γ elements.

Write these as $J_I \backslash I' = \{j'_1, j'_2, \ldots, j'_\gamma\}$ with $j'_1 < j'_2 < \cdots < j'_\gamma$. Define the map $\Phi_{\varphi,I,I'}$: $M \times N \to \mathbb{R}^\gamma$ by

$$\Phi_{\varphi,I,I'}(x, y) = (\varphi(x_{j'_1}), \varphi(x_{j'_2}), \ldots, \varphi(x_{j'_\gamma}))^\dagger.$$

Observe that if for some $v \in \tilde{T}_{I,I'}$ we have $T_{(x,y)}\Phi_{\varphi,I,I'}(v) \neq 0$, then $T_{(x,y)}\Phi_{f,g,\varphi}(v) \neq 0$. Thus if $T_{(x,y)}\Phi_{\varphi,I,I'}(v) \neq 0$ for all $v \in \tilde{T}_{I,I'}$ and all I and I', then $\Phi_{f,g,\varphi}$ is immersive at (x, y). Thus, to show the immersivity of $\Phi_{f,g,\varphi}$ on the whole of $(M \times N)\backslash\mathcal{P}_{f,g}$, it is sufficient to show that the image of each $\tilde{T}_{I,I'}$ under $T\Phi_{\varphi,I,I'}$ does not intersect the zero section in $T\mathbb{R}^\gamma$.

Define the map ρ: $\mathcal{C}^r(M, \mathbb{R}) \to \mathcal{C}^1(\tilde{T}_{I,I'}, T\mathbb{R}^\gamma)$ by

$$\rho(\varphi) = \tilde{T}\Phi_{\varphi,I,I'}.$$

Let $L_{I,I'}$ be the zero section in $T\mathbb{R}^\gamma$. In Section 5.5 below we prove

Proposition 5.6. *The map ev_ρ is transversal to $L_{I,I'}$ for all I and I'.*

By Corollary C.5, ev_ρ is $\mathcal{C}^1$ if $r \geq 3$. The dimension of $\tilde{T}_{I,I'}$ is $2m+2n-1-(d-\gamma)m$ and the codimension of $L_{I,I'}$ is γ, and thus

$$\begin{aligned}\dim \tilde{T}_{I,I'} - \operatorname{codim} L_{I,I'} &= 2m + 2n - 1 - (d-\gamma)m - \gamma \\ &\leq d - 2 - (d-\gamma)m - \gamma \\ &\leq (d-\gamma)(1-m) - 2 \\ &< 0,\end{aligned}$$

since $d \geq \gamma$ and $m \geq 1$. Hence by the Parametric Transversality Theorem, there is a residual set of φ for which $\tilde{T}\Phi_{\varphi,I,I'}$ is transversal to $L_{I,I'}$. Now, just as in the proof of the standard Takens Theorem, we count dimensions: If $v \in \tilde{T}_{I,I'}$ then $\dim T_v(\tilde{T}_{I,I'}) = 2m + 2n-1-(d-\gamma)m$ and hence the dimension of its image is at most $2m+2n-1-(d-\gamma)m$, whilst the dimension of $T_u L_{I,I'}$ is γ for any $u \in L_{I,I'}$. Hence if $u = \tilde{T}\Phi_{\varphi,I,I'}(v) = 0$, then $T_v(\tilde{T}\Phi_{\varphi,I,I'})(T_v(\tilde{T}_{I,I'}))+T_u L_{I,I'}$ has dimension at most $2m+2n-1-(d-\gamma)m+\gamma$. But from the inequality above, we have $2m + 2n - 1 - (d-\gamma)m + \gamma < 2\gamma$, and hence $T_v(\tilde{T}\Phi_{\varphi,I,I'})(T_v(\tilde{T}_{I,I'}))+T_u L_{I,I'}$ cannot span $T_u(T\mathbb{R}^\gamma)$. Thus if $\tilde{T}\Phi_{\varphi,I,I'}$ is transversal to $L_{I,I'}$, then its image cannot intersect $L_{I,I'}$. Hence for a residual set of φ, $\tilde{T}\Phi_{\varphi,I,I'}(v) \neq 0$ for all $v \in \tilde{T}_{I,I'}$. Taking the intersection over all I, I', we get a residual set of φ such that $\tilde{T}\Phi_{f,g,\varphi}(v) \neq 0$ for all $v \in ((M \times N)\backslash\mathcal{P}_{f,g})$, and hence $\Phi_{f,g,\varphi}$ is immersive as required.

5.1.4. Injectivity. To prove the injectivity of $\Phi_{f,g,\varphi}$ for a residual set, we want to construct a set of manifolds $\tilde{W}_{I,R} \subset \tilde{W}_I \times (M \times N)$ that play an analogous role to the $\tilde{T}_{I,I'}$ above. Here R is a subset of $J_I \times J_I$ and we want points in $\tilde{W}_{I,R}$ to satisfy $x_i = x'_{i'}$ for all $(i, i') \in R$ (where as usual $x_i = f^{(i)}(x, y)$ and $x'_i = f^{(i)}(x', y')$). Unfortunately when we attempt to construct $\tilde{W}_{I,R}$ using a similar argument to that for $\tilde{T}_{I,I'}$, we run into difficulties at pairs of points (x, y), (x', y') such that either (x', y') is periodic of period less than d, or such that $(x', y') = (f, g)^k(x, y)$ for some k with $-d \leq k \leq d$.

In particular, the appropriate evaluation map can fail to be transversal at such points. This lack of transversality is actually to be expected for essentially the same reasons as discussed above in the context of W_I and the difficulties played by periodic orbits in its construction. Thus, for instance, if $(x', y') = (f, g)^k(x, y)$, then $x_i + k = x'_i$ for all i. Hence if $(j + k, j) \in R$ for some j, then $(i + k, i) \in R$ for all i. The relations in R are therefore not independent in this case, leading to a loss of transversality. We therefore need to treat such points separately and will define two other families of submanifolds $\tilde{W}_{\tilde{I},(k)}$ and $\tilde{W}_{I,R',(q)}$ in which they lie. The treatment of these is very similar to that of $\tilde{W}_{I,R}$, but there appears to be no convenient way of unifying all three cases. Furthermore, whilst in the case of $\tilde{W}_{I,R}$ and $\tilde{W}_{I,R',(q)}$ it is sufficient to exclude periodic orbits of period less than d, in the case of $\tilde{W}_{\tilde{I},(k)}$, we need to exclude all periods up to $d + k$. This was the reason why we gave a separate proof of injectivity for $\mathcal{P}^{(2d)}_{f,g} \times \mathcal{P}^{(2d)}_{f,g}$ rather than just for $\mathcal{P}_{f,g} \times \mathcal{P}_{f,g}$.

First, we construct $\tilde{W}_{I,R}$. Let $R = \{(i_1, i'_1), (i_2, i'_2), \ldots, (i_{\beta_R}, i'_{\beta_R})\}$ be a subset of $J_I \times J_I$ (possibly empty) and define

$$\begin{aligned}\tilde{W}_{I,R} = \{&((x, y), (x', y')) \in \tilde{W}_I \times (M \times N)\backslash\mathcal{P}_{f,g},\\ &(x', y') \neq (f, g)^k(x, y) \text{ for any } k \text{ such that } -d \leq k \leq d \text{ and}\\ &\text{if } (i, i') \in J_I \times J_I \text{ then } x_i = x'_{i'} \text{ if and only if } (i, i') \in R\}.\end{aligned}$$

Note that this definition is independent of the choice of J_I. Also observe that for some choices of R, $\tilde{W}_{I,R}$ must necessarily be empty. Thus, for instance, if $(i, i') \in R$ and $(j, i') \in R$ for some $i \neq j$ we must have $x_i = x'_{i'} = x_j$, for all $((x, y), (x', y')) \in \tilde{W}_{I,R}$. On the other hand, by the definition of J_I, $x_i \neq x_j$ for all $i, j \in J_I$ such that $i \neq j$, and hence $\tilde{W}_{I,R} = \varnothing$. However, since we cannot guarantee for any R that $\tilde{W}_{I,R}$ is nonempty for a residual set of f we do not bother to exclude those R for which $\tilde{W}_{I,R}$ is always empty. Let $\alpha_R = \alpha - \beta_R$. A similar proof to that of Proposition 5.5, given in Section 5.6 below, leads to

Proposition 5.7. *For a residual set of $f \in \mathcal{D}^r$, if $\tilde{W}_{I,R}$ is nonempty then it is a codimension $(d - \alpha_R)m$ submanifold of $(M \times N)\backslash\mathcal{P}_{f,g} \times (M \times N)\backslash\mathcal{P}_{f,g}$.*

Next, we turn to the case where one of (x, y) or (x', y') is periodic. By interchanging (x, y) and (x', y') if necessary, it suffices to restrict attention to $(M \times N)\backslash\mathcal{P}_{f,g} \times \mathcal{P}_{f,g}$. Let $R' = \{(i_1, i'_1), (i_2, i'_2), \ldots, (i_{\beta_{R'}}, i'_{\beta_{R'}})\}$ be a subset of $J_I \times \{0, \ldots, q - 1\}$ (possibly empty) and for $1 \leq q < 2d$ define

$$\begin{aligned}\tilde{W}_{I,R',(q)} = \{&((x, y), (x', y')) \in \tilde{W}_I \times (M \times N),\\ &(x', y') \text{ is periodic of minimal period } q \text{ and}\\ &\text{if } (i, i') \in J_I \times 0, \ldots, q - 1 \text{ then } x_i = x'_{i'} \text{ if and only if } (i, i') \in R'\},\end{aligned}$$

As before, we do not bother to exclude the cases where $\tilde{W}_{I,R',(q)}$ has to be empty. Let $\alpha_{R'} = \alpha - \beta_{R'}$. A similar argument to the proof of Proposition 5.6 (found in Section 5.7) gives the following.

Proposition 5.8. *For a residual set of* $f \in \mathcal{D}^r$, *if* $\tilde{W}_{I,R',(q)}$ *is nonempty, then it is a codimension* $(d - \alpha_{R'})m + m + n$ *submanifold of* $(M \times N)\backslash\mathcal{P}_{f,g} \times (M \times N)$.

The additional $m + n$ in the codimension arises from the fact that periodic orbits of period q are isolated, and thus have codimension $m + n$ in $M \times N$.

Finally, we treat points such that $(x', y') = (f, g)^k(x, y)$ for some $k \neq 0$ such that $-d \leq k \leq d$. Without loss of generality (interchanging (x, y) and (x', y') if necessary), we may assume that $k > 0$. Let $\tilde{I} = \{\tilde{I}_1, \tilde{I}_2, \ldots, \tilde{I}_{\tilde{\alpha}}\}$ be a partition of $\{0, \ldots, d + k - 1\}$. Observe that the only condition on d in Proposition 5.4 is that $\tilde{W}_I$ does not contain periodic orbits of period $< d$. Hence this proposition implies that

$$\begin{aligned}\tilde{W}_{\tilde{I}} = \{&(x, y) \in (M \times N)\backslash\mathcal{P}^{(d+k)} f, g\colon\ x_i = x_{i'} \text{ if and only if}\\ &i, i' \text{ are in the same element of } \tilde{I}\}\end{aligned}$$

for a residual set of $f \in \mathcal{D}^r$ is either empty, or a codimension $(d + k - \tilde{\alpha})m$ submanifold of $(M \times N)\backslash\mathcal{P}^{(d+k)} f, g$. We then define

$$\tilde{W}_{\tilde{I},(k)} = \{(x, y), (f, g)^k(x, y)\colon\ (x, y) \in \tilde{W}_{\tilde{I}}\}.$$

Since (f, g) is a diffeomorphism, this is a submanifold of $(M \times N)\backslash\mathcal{P}^{(d+k)} f, g \times (M \times N)\backslash\mathcal{P}^{(d+k)} f, g$ whenever $\tilde{W}_{\tilde{I}}$ is. Its dimension is then the same as that of $\tilde{W}_{\tilde{I}}$, and hence its codimension is $(d + k - \tilde{\alpha})m + m + n$.

Since the definitions of the various $\tilde{W}$ are somewhat intricate, we give a detailed argument to show that the union of $\mathcal{P}^{(2d)}_{f,g} \times \mathcal{P}^{(2d)}_{f,g}$ and the sets $\tilde{W}_{I,R}$, $\tilde{W}_{I,R',(q)}$, and $\tilde{W}_{\tilde{I}(k)}$ over all I, R, R', k, and q, includes the whole of $((M \times N) \times (M \times N))\backslash\Delta$, up to an interchange of (x, y) and (x', y') if necessary. Thus consider an arbitrary$((x, y), (x', y')) \in ((M \times N) \times (M \times N))\backslash\Delta$. If both (x, y) and (x', y') are periodic of period $< 2d$, then $((x, y), (x', y')) \in \mathcal{P}^{(2d)}_{f,g} \times \mathcal{P}^{(2d)}_{f,g}$. If only one is periodic, say without loss of generality (x', y'), then if the period $q < d$, then $((x, y), (x', y')) \in \tilde{W}_{I,R',(q)}$, with I determined by the relations amongst $\{x_0, \ldots, x_{d-1}\}$ and R' by the relations between $\{x_0, \ldots, x_{d-1}\}$ and $\{x'_0, \ldots, x'_{q-1}\}$. If the period q of (x', y') is such that $d \leq q < 2d$, then $((x, y), (x', y')) \in \tilde{W}_{I,R}$ with I and R determined similarly. If neither (x, y) and (x', y') are periodic of period $< 2d$, then if $(x', y') = (f, g)^k(x, y)$ for some $k \neq 0$ such that $-d \leq k \leq d$ we have either $((x, y), (x', y')) \in \tilde{W}_{\tilde{I},(k)}$ or $((x', y'), (x, y)) \in \tilde{W}_{\tilde{I},(-k)}$ for an appropriate $\tilde{I}$, and if $(x', y') \neq (f, g)^k(x, y)$, then $((x, y), (x', y')) \in \tilde{W}_{I,R}$.

By taking a finite intersection we get a residual set of $f \in \mathcal{D}^r$ such that all the $\tilde{W}$ are submanifolds. Fix one such f. For each of the $\tilde{W}_{I,R}$, $\tilde{W}_{I,R',(q)}$, and $\tilde{W}_{\tilde{I},(k)}$, we shall now show that there is a residual set of $\varphi \in \mathcal{A}_{f,g}$ such that their image under $\Phi_{f,g,\varphi} \times \Phi_{f,g,\varphi}$ does not intersect the diagonal in $\mathbb{R}^d \times \mathbb{R}^d$. As usual, we will do this by showing transversality for an appropriate evaluation map, followed by counting dimensions.

First let us consider $\tilde{W}_{I,R}$. Fix a particular I and R and recall that we have chosen an f such that $\tilde{W}_{I,R}$ is a codimension $(d - \alpha_R)m$ submanifold, where $\alpha_R = \alpha - \beta_R$. Note that we only need consider the case $\beta_R < \alpha$, for if $\beta_R \geq \alpha$ and $\tilde{W}_{I,R}$ were nonempty, then its dimension would be $2m + 2n - (d - (\alpha - \beta_R))m \leq 2m + 2n - dm \leq d - 1 - dm \leq d(1 - m) - 1 < 0$. Hence if we define

$$J_{I,R} = \{i \in J_I\colon\ (i, i') \notin R \text{ for any } i' \in J_I\},$$

we see that $J_{I,R}$ is nonempty. Let γ' be the number of elements in $J_{I,R}$ and note that $\gamma' \geq \alpha - \beta_R$. Write $J_{I,R} = \{j'_1, j'_2, \ldots, j'_{\gamma'}\}$, with $j'_1 < j'_2 < \cdots < j'_{\gamma'}$. Define the map $\Phi_{\varphi,I,R}\colon M \times N \to \mathbb{R}^{\gamma'}$ by

$$\Phi_{\varphi,I,R}(x, y) = (\varphi(x_{j'_1}), \varphi(x_{j'_2}), \ldots, \varphi(x_{j'_{\gamma'}}))^{\dagger}.$$

If $\Phi_{\varphi,I,R}(x, y) \neq \Phi_{\varphi,I,R}(x', y')$, then $\Phi_{f,g,\varphi}(x, y) \neq \Phi_{f,g,\varphi}(x', y')$. Define the map $\sigma\colon \mathcal{C}^r(M, \mathbb{R}) \to \mathcal{C}^1(\tilde{W}_{I,R}, \mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'})$ by

$$\sigma(\varphi)(x, y, x', y') = (\Phi_{\varphi,I,R}(x, y), \Phi_{\varphi,I,R}(x', y')).$$

Since for $(x, y, x', y') \in \tilde{W}_{I,R}$, the points $x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}$ are distinct and disjoint from $x'_{j'_1}, x'_{j'_2}, \ldots, x'_{j'_{\gamma'}}$, it is straightforward to show the following.

Proposition 5.9. *The map ev_σ is transversal to the diagonal $\hat{\Delta}_{\gamma'}$ in $\mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'}$ for all I and R.*

The proof appears in Section 5.8 below. By Corollary C.3, ev_σ is $\mathcal{C}^1$ if $r \geq 2$. The dimension of $\tilde{W}_{I,R}$ is $2m + 2n - (d - a_R)m$ where $\alpha_R = \alpha - \beta_R$ and the codimension of $\hat{\Delta}_{\gamma'}$ is γ'. Thus, since $\gamma' \geq \alpha_R$,

$$\begin{aligned} \dim \tilde{W}_{I,R} - \operatorname{codim} \hat{\Delta}_{\gamma''} &= 2m + 2n - (d - \alpha_R)m - \gamma' \\ &\leq d - (d - \alpha_R)m - \gamma' - 1 \\ &\leq (d - \alpha_R)(1 - m) - 1 \\ &< 0. \end{aligned}$$

Hence, by the Parametric Transversality Theorem, there is a residual set of φ for which $\sigma(\varphi)$ is transversal to $\hat{\Delta}_{\gamma'}$. Now, once again, we just count dimensions: The dimension of $T_{x,y,x',y'}(\tilde{W}_{I,R})$ is $2m + 2n - (d - \alpha_R)m$ and hence the dimension of its image is less than or equal to $2m + 2n - (d - \alpha_R)m$. The dimension of $T_{z,z}\hat{\Delta}_{\gamma'}$ is γ' for any $(z, z) \in \hat{\Delta}_{\gamma'}$. Thus if $\Phi_{\varphi,I,R}(x, y) = \Phi_{\varphi,I,R}(x', y')$ so that $\sigma(\varphi)(x, y, x', y') \in \hat{\Delta}_{\gamma'}$, then $T_{x,y,x',y'}\sigma(\varphi)(T_{x,y,x',y'}(\tilde{W}_{I,R})) + T_{z,z}\hat{\Delta}_{\gamma'}$ has dimension at most $2m + 2n - (d - \alpha_R)m + \gamma'$. From the inequality above $2m + 2n - (d - \alpha_R)m + \gamma' < 2\gamma'$, and hence $T_{x,y,x',y'}\sigma(\varphi)(T_{x,y,x',y'}(\tilde{W}_{I,R})) + T_{z,z}\hat{\Delta}_{\gamma'}$ cannot span $T_{z,z}\mathbb{R}^{\gamma'}$. Thus if $\sigma(\varphi)$ is transversal to $\hat{\Delta}_{\gamma'}$, then it cannot intersect $\hat{\Delta}_{\gamma'}$. Hence for a residual set of φ we have $\Phi_{\varphi,I,R}(x, y) \neq \Phi_{\varphi,I,R}(x', y')$ for all $(x, y, x', y') \in \tilde{W}_{I,R}$.

Next, we turn to $\tilde{W}_{I,R',(q)}$. The argument is very similar to that for $\tilde{W}_{I,R}$. Thus let

$$J'_{I,R} = \{i \in J_I\colon (i, i') \notin R' \text{ for any } i' \in J_I\}.$$

As before, without loss of generality we can restrict ourselves to $\beta_{R'} < \alpha$, and so assume that $J'_{I,R}$ is nonempty. Let γ' be the number of elements in $J'_{I,R}$ and note that $\gamma' \geq \alpha - \beta_R$. Write $J'_{I,R} = \{j'_1, j'_2, \ldots, j'_{\gamma'}\}$, with $j'_1 < j'_2 < \cdots < j'_{\gamma'}$, and define the map $\Phi_{\varphi,I,R'}\colon M \times N \to \mathbb{R}^{\gamma'}$ by

$$\Phi_{\varphi,I,R'}(x, y) = (\varphi(x_{j'_1}), \varphi(x_{j'2}), \ldots, \varphi(x_{j'_{\gamma'}}))^{\dagger},$$

and $\sigma'\colon \mathcal{C}^r(M,\mathbb{R}) \to \mathcal{C}^r(\tilde{W}_{I,R'},(q),\mathbb{R}^{\gamma'}\times\mathbb{R}^{\gamma'})$ by

$$\sigma'(\varphi)(x,y,x',y') = (\Phi_{\varphi,I,R'}(x,y),\Phi_{\varphi,I,R'}(x',y')).$$

If $(x,y,x',y') \in \tilde{W}_{I,R',(q)}$, then the points $x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}$ are distinct and $x_{j'_i} \neq x'_j$ for all $j'_i \in J'_{I,R}$ and $j \in \{0,\ldots,d-1\}$. The latter follows from the fact that $x_{j'_i} \neq x'_j$ for all $j \in \{0,\ldots,q-1\}$, and since (x',y') has period q, $x'_j = x'_j \mod q$ for all $j \in \{0,\ldots,d-1\}$. Hence we proceed exactly as above, and in Section 5.9 below we show the following.

Proposition 5.10. *The map ev'_σ is transversal to the diagonal $\hat{\Delta}\gamma'$ in $\mathbb{R}^{\gamma'}\times\mathbb{R}^{\gamma'}$ for all I, R', and q.*

By Corollary C.3, ev_σ is $\mathcal{C}^1$ if $r \geq 2$. The dimension of $\tilde{W}_{I,R',(q)}$ is $m+n-(d-\alpha_{R'})m$. Thus, since $\gamma' \geq \alpha_{R'}$,

$$\begin{aligned}
\dim \tilde{W}_{I,R} - \operatorname{codim}\hat{\Delta}_{\gamma''} &= m+n-(d-\alpha_{R'})m-\gamma' \\
&< d-(d-\alpha_{R'})m-\gamma'-m-n \\
&\leq (d-\alpha_{R'})(1-m)-m-n \\
&< 0.
\end{aligned}$$

Hence, by the Parametric Transversality Theorem there is a residual set of φ for which $\sigma'(\varphi)$ is transversal to $\hat{\Delta}\gamma'$. Counting dimensions, the dimension of $T_{z,z}\hat{\Delta}_{\gamma'}+T_{x,y,x',y'}\sigma(\varphi)(T_{x,y,x',y'}(\tilde{W}_{I,R',(q)}))$ is at most $m+n-(d-\alpha_{R'})m+\gamma'$ which from the inequality above is strictly less than $2\gamma'$. Hence if $\sigma(\varphi)$ is transversal to $\hat{\Delta}_{\gamma'}$, it cannot intersect $\hat{\Delta}_{\gamma'}$ and so we conclude that for a residual set of φ we have $\Phi_{\varphi,I,R'}(x,y) \neq \Phi_{\varphi,I,R'}(x',y')$, and thus $\Phi_{f,g,\varphi}(x,y) \neq \Phi_{f,g,\varphi}(x',y')$ for all $(x,y,x',y') \in \tilde{W}_{I,R',(q)}$.

Finally, we deal with $\tilde{W}_{\tilde{I},(k)}$. Recall that $\tilde{I} = \{\tilde{I}_1,\tilde{I}_2,\ldots,\tilde{I}_{\tilde{\alpha}}\}$ is a partition of $\{0,\ldots,d+k-1\}$. Let $\tilde{I}'$ be a subset of $\tilde{I}$ defined by

$$\tilde{I}' = \{\tilde{I}_i \in \tilde{I}\colon\ \tilde{I}_i \cap \{d,\ldots,d+k-1\} = \varnothing\}.$$

Let $J_{\tilde{I}}$ be a set containing the largest element j'_i from each $\tilde{I}_i \in \tilde{I}$. Re-order $\tilde{I}'$ so that $j'_1 < j'_2 < \cdots < j_{\tilde{\gamma}}$, where $\tilde{\gamma}$ is the number of elements in $\tilde{I}'$. Note that in forming $\tilde{I}'$ we exclude at most k elements of $\tilde{I}$, and hence $\tilde{\gamma} \geq \tilde{\alpha}-k$. Hence if $\tilde{I}' = \varnothing$, we must have $\tilde{\alpha}-k \leq 0$ and if $\tilde{W}_{\tilde{I},(k)}$ were nonempty, its dimension would be $2m+2n-((d+k-\alpha)m+m+n) \leq m+n-dm < d(1-m)-m-n < 0$. Hence without loss of generality we can restrict ourselves to those $\tilde{I}$ such that $\tilde{\gamma} > 0$. We then define $\Phi_{\varphi,\tilde{I}'}\colon M\times N \to \mathbb{R}^{\tilde{\gamma}}$ by

$$\Phi_{\varphi,\tilde{I}'}(x,y) = (\varphi(x_{j'_1}),\varphi(x_{j'_2}),\ldots,\varphi(x_{j'_{\tilde{\gamma}}}))^\dagger,$$

and $\sigma''\colon \mathcal{C}^r(M,\mathbb{R}) \to \mathcal{C}^r(\tilde{W}_{\tilde{I},(k)},\mathbb{R}^{\tilde{\gamma}}\times\mathbb{R}^{\tilde{\gamma}})$

$$\sigma''(\varphi)(x,y,x',y') = (\Phi_{\varphi,\tilde{I}'}(x,y),\Phi_{\varphi,\tilde{I}'}(x',y')).$$

A slightly more delicate argument than the previous two propositions then yields

Proposition 5.11. *The map $ev_{\sigma''}$ is transversal to the diagonal $\hat{\Delta}_{\tilde{\gamma}}$ in $\mathbb{R}^{\tilde{\gamma}} \times \mathbb{R}^{\tilde{\gamma}}$ for all $\tilde{I}$ and k.*

The proof is given in Section 5.10 below. By Corollary C.3, $ev_{\sigma'}$ is $\mathcal{C}^1$ if $r \geq 2$. The dimensions of $\tilde{W}_{\tilde{I},(k)}$ and $\hat{\Delta}_{\tilde{\gamma}}$ are respectively $2m + 2n - ((d + k - \tilde{\alpha})m + m + n)$ and $\tilde{\gamma}$, and since $\tilde{\alpha} - k \leq \tilde{\gamma} \leq d$, the usual computation gives

$$\begin{aligned} \dim \tilde{W}_{\tilde{I},(k)} - \operatorname{codim} \hat{\Delta}_{\alpha} &= 2m + 2n - ((d + k - \tilde{\alpha})m + m + n) - \tilde{\gamma} \\ &< d - dm + (\tilde{\alpha} - k)m - \tilde{\gamma} - m - n \\ &\leq (d - \tilde{\gamma})(1 - m) - m - n \\ &< 0. \end{aligned}$$

Hence, by the Parametric Transversality Theorem there is a residual set of φ for which $\sigma''(\varphi)$ is transversal to $\hat{\Delta}_{\tilde{\gamma}}$. Then, as above, the dimension of $T_{x,y,x',y'}(\tilde{W}_{\tilde{I},(k)})$ is $m+n-((d+k-\tilde{\alpha})m)$ and the dimension of $T_{z,z}\hat{\Delta}_{\tilde{\gamma}}$ is $\tilde{\gamma}$. Since $m+n-((d+k-\tilde{\alpha})m)+\tilde{\gamma} < 2\tilde{\gamma}$, the only way that $\sigma''(\varphi)$ can be transversal to $\hat{\Delta}_{\tilde{\gamma}}$ is by not intersecting $\hat{\Delta}_{\tilde{\gamma}}$. Hence for a residual set of φ, we have $\Phi_{\varphi,\tilde{I}}(x, y) \neq \Phi_{\varphi,\tilde{I}'}(x', y')$ for all $(x, y, x', y') \in \tilde{W}_{\tilde{I},(k)}$, and thus $\Phi_{f,g,\varphi}(x, y) \neq \Phi_{f,g,\varphi}(x', y')$ for all $(x, y, x', y') \in \tilde{W}_{\tilde{I},(k)}$.

This concludes the proof of Theorem 3.1.

5.2. Embedding the Short Periodic Points

5.2.1. Isolating the Periodic Points.

Proof of Proposition 5.1. This follows very similar lines to the proof of Proposition 4.1, though it is technically somewhat messier. Our first task, as before, is to show that for a dense open set of f, the periodic orbits of any given period are isolated and hyperbolic. If (x, y) is periodic under (f, g), then obviously y is periodic under g. By hypothesis, such y are isolated. Fix one such y, and for $q \geq 1$, let $\mathcal{B}_{y,q}$ be the set of $f \in \mathcal{D}^r(M \times N, M)$ for which for all $k \geq 1$, $k < q$, the set of x such that (x, y) is periodic of period k is a finite set of isolated points, and such that these points are hyperbolic.

Denote by $T_{1,(x,y)} f^{(k)}$ and $T_{2,(x,y)} f^{(k)}$ the "partial derivatives" of $f^{(k)}$, i.e., by definition $T_{1,(x,y)} f^{(k)}(v) = T_{(x,y)} f^{(k)}(v, 0)$ and $T_{2,(x,y)} f^{(k)}(u) = T_{(x,y)} f^{(k)}(0, u)$ for $v \in T_x M$ and $u \in T_y N$. Then the eigenvalues of $T_{(x,y)}(f, g)^k$ are the union of the eigenvalues of $T_{1,(x,y)} f^{(k)}$ and $T_y g^k$. Since by hypothesis y is a hyperbolic periodic orbit of g, to show that (x, y) is hyperbolic it is thus sufficient to show that $T_{1,(x,y)} f^{(k)}$ has no eigenvalues that are a root of unity. By default we define $\mathcal{B}_{y,1} = \mathcal{D}^r(M \times N, M)$.

Our inductive hypothesis is that $\mathcal{B}_{y,q}$ is open and dense in $\mathcal{D}^r(M \times N, M)$. This holds trivially for $q = 1$, so now suppose that it holds for some $q > 1$. We shall apply the Parametric Transversality Theorem with $\rho\colon \mathcal{B}_{y,q} \to \mathcal{C}^r(M \times \{y\}, M \times M)$ given by $\rho(f) = (f^{(q)}, f^{(0)})$, so that the evaluation map is $ev_\rho(f, x, y) = (f^{(q)}(x, y), x)$. Note that ρ is thus just two components of the map τ in the previous section. Let $\Delta = \{(x, x)\colon x \in M\}$ be the diagonal in $M \times M$; we aim to show that ev_ρ is transversal to it.

Let $T_{f,x,y}(ev_\rho)(\eta, 0) = (uq, 0)$. Thus by Corollary C.5, we have

$$\begin{aligned} u_q &= \sum_{i=1}^{q} T_{(x_i,y_i)} f^{(q-i)}(\eta(x_{i-1}, y_{i-1}), 0) \\ &= \sum_{i=1}^{p} T^{(k)}(T_{(x_i,y_i)} f^{(p-i)}(\eta(x_{i-1}, y_{i-1}), 0), \end{aligned}$$

where p is the least period of (x, y), $T_{(x_j,y_j)} f^{(0)}(u, 0) = u$ and

$$T^{(k)} = \sum_{j=1}^{k} (T_{1,(x,y)} f^{(p)})^{j-1},$$

where $q = kp$. If $k = 1$, $T^{(k)} = Id$ and hence is invertible, otherwise (x, y) is periodic of period $p < q$, and hence by our inductive hypothesis none of the eigenvalues of $T_{1,(x,y)} f^{(p)}$ are roots of unity. The eigenvalues of $T^{(k)}$ are of the form $s(\lambda) = 1 + \lambda + \cdots + \lambda^{k-1}$ for λ an eigenvalue of $T_{1,(x,y)} f^{(p)}$, and $s(\lambda)(1-\lambda) = (1-\lambda^k)$. Thus $s(\lambda) \neq 0$ for λ an eigenvalue of $T_{1,(x,y)} f^{(p)}$, and so the eigenvalues of $T^{(k)}$ are nonzero. Hence $T^{(k)}$ is invertible, and given any $v \in T_x M$, we can choose $\eta \in T_f \mathcal{B}_q$ such that $\eta(x_{p-1}, y_{p-1}) = (T^{(k)})^{-1}(v)$ and $\eta(x_i, y_i) = 0$ for $i = 0, \ldots, p-2$. For such a η, we have $T_{f,x,y}(ev_\rho)(\eta, 0) = (v, 0)$. Since $(v, v) \in T_{x,x}\Delta$, we have $(0, v) \in T_{x,x}\Delta +$ Image $T_{f,x,y}(ev_\rho)$. Together the $(v, 0)$ and $(0, v)$ generate $T_x M \times T_x M$ and hence ev_ρ is transversal to Δ, as required.

Thus, by the Parametric Transversality Theorem, there is an open and dense subset of $\mathcal{B}_{y,q}$ such that $(f^{(q)}, f^{(0)})$ is transversal to Δ. For such maps the set $(f^{(q)}, f^{(0)})^{-1}(\Delta)$ is a codimension m submanifold of $M \times \{y\}$, and hence is a finite number of isolated points. But $(f^{(q)}, f^{(0)})^{-1}(\Delta)$ is precisely the set of periodic orbits of period q in $M \times \{y\}$. To complete the inductive step we need to show that we can further restrict the set of f to ensure that $T_{1,(x,y)} f^{(q)}$ has no eigenvalues that are a root of unity; this is done in exactly the same way as in Proposition 4.1.

We have thus shown that $\mathcal{B}_{y,q+1}$ is open and dense in $\mathcal{B}_{y,q}$, and since $\mathcal{B}_{y,q}$ is open and dense in $\mathcal{D}^r(M \times N, M)$, $\mathcal{B}_{y,q+1}$ is also open and dense in $\mathcal{D}^r(M \times N, M)$, thus completing the inductive step. Now, by taking the intersection of the $\mathcal{B}_{y,q}$ over y a periodic orbit of g, and $q < 2d$, we obtain an open and dense subset $\mathcal{B}_{2d}$ of $\mathcal{D}^r(M \times N, M)$ which has only a finite number of periodic orbits of period $q < 2d$, and such that these are all hyperbolic. Again as in Proposition 4.1 we can easily restrict ourselves to an open and dense subset $\mathcal{B}'_{2d}$ of $\mathcal{B}_{2d}$ for which the eigenvalues of any such periodic orbit are all distinct. Also, by the Transversal Isotopy Theorem, sufficiently small perturbations to $f \in \mathcal{B}'_{2d}$ do not change the number of periodic orbits of each period less than $2d$.

Finally, to obtain $\mathcal{D}^r$ it remains to restrict $\mathcal{B}'_{2d}$ further to ensure that x coordinates of any two distinct periodic points of periods less than $2d$ are different. Observe that if $f \in \mathcal{B}'_{2d}$ and x is a periodic point of period less than $2d$, then by the Implicit Function Theorem, x depends smoothly on f. Since by the above small perturbations of f do not change the number of such periodic orbits, we see that $\mathcal{D}^r$ is open in $\mathcal{B}'_{2d}$. Conversely if $f \in \mathcal{B}'_{2d}$, then using Corollary C.13 we can move the x coordinates of the periodic points of f as follows: Suppose that $f(x_{i-1}, y_{i-1}) = x_i$ and $f(x_i, y_i) = x_{i+1}$, with

$(x_{i-1}, y_{i-1}) \neq (x_i, y_i)$. Then by Corollary C.13, if x_i' is sufficiently close to x_i we can find an f' such that $f'(x_{i-1}, y_{i-1}) = x_i'$, $f'(x_i', y_i) = x_{i+1}$ and $f' = f$ at all other periodic points of f. Thus f' has the same periodic orbits as f except that we have shifted x_i to x_i'. This argument fails if $(x_i, y_i) = (x, y)$ is a fixed point, but in that case we can still choose f' close to f such that $f'(x, y) \neq x$, and hence (x, y) is no longer a fixed point of f'. Hence if f had two (or more) fixed points (x, y) and (x, y') with the same first coordinate, we can ensure that (x, y) is the only fixed point of f' whose first coordinate is x. Given any $f \in \mathcal{B}_{2d}'$ we therefore can find an arbitrarily close f' all of whose periodic points of period less than $2d$ have distinct first coordinates, and thus in other words $\mathcal{D}^r$ is dense in $\mathcal{B}_{2d}'$. □

5.2.2. Embedding the Periodic Points.

Proof of Proposition 5.2. Recall that a sufficient condition for $T_{x,y}\Phi_{f,g,\varphi}$ to be 1-1 is for the first $m+n$ components of $T_{x,y}\Phi_{f,g,\varphi}$ to be linearly independent. The i^{th} such component is given by $a_i = T_{x_i}\varphi \circ T_{(x,y)} f^{(i)} \in \mathcal{L}(T_x M \times T_y N, T_{\varphi(x_i)}\mathbb{R})$. Note that we may write this as $a_i = T_{x_i}\varphi \circ T_{(x_i,y_i)}\pi_1 \circ T_{(x,y)}(f, g)^i$ and thus in effect we are restricted to observation functions $\Psi\colon M \times N \to \mathbb{R}$ such that $T_{(x_i,y_i)}\Psi$ lies in the subset of $\mathcal{L}(T_x M \times T_y N, \mathbb{R})$ of the form $T_{x_i}\varphi \circ T_{(x_i,y_i)}\pi_1$. The only way that $T_{x_i}\varphi$ can "observe" $T_y N$ is thus via $T_{2,(x,y)} f^{(i)}$, in particular $T_{x_i}\varphi \circ T_{(x,y)} f^{(i)}(v, u) = T_{x_i}\varphi(T_{1,(x,y)} f^{(i)}(v, 0) + T_{2,(x,y)} f^{(i)}(0, u))$. In effect therefore, the proof below shows that we can make $T_{x_i}\varphi \circ T_{2,(x,y)} f^{(i)}$ be whatever we want by an appropriate choice of $T_{2,(x_i,y_i)} f$ at each (x_i, y_i).

Suppose (x, y) has minimal period q. Since $(x_0, y_0), \ldots, (x_{q-1}, y_{q-1})$ are distinct, the maps $T_{x_0}\varphi, \ldots, T_{x_{q-1}}\varphi$ can be adjusted independently, and similarly for $T_{2,(x_0,y_0)} f$, $\ldots, T_{2,(x_{q-1},y_{q-1})} f$. As in Section 4.2, to show that the first $m+n$ components of $T_{x,y}\Phi_{f,g,\varphi}$ are linearly independent we need to show that if $\omega_a = a_0 \wedge \cdots \wedge a_{m+n-1}$ then $\omega_a(e_0, \ldots, e_{m+n1}) \neq 0$ for some basis of $T_x M \times T_y N$. Let $\alpha^{(i)} \in \mathbb{R}^m$, $A^{(i)} \in \mathbb{R}^{m\times m}$, and $B^{(i)} \in \mathbb{R}^{m\times n}$ for $i = 0, \ldots, q-1$, be the matrices representing $T_{x_i}\varphi$, $T_{1,(x_i,y_i)} f$ and $T_{2,(x_i,y_i)} f$ with respect to the bases $v_0^{(i)}, \ldots, v_{m-1}^{(i)}$ and $u_m^{(i)}, \ldots, u_{m+n-1}^{(i)}$ of $T_{x_i} M$ and $T_{y_i} N$, respectively, where $(v_j^{(i)}, u_j^{(i)}) = T_{(x,y)}(f, g)^i(e_j)$. As we shall see for a particular choice of basis (and hence for any basis) and any given $A^{(i)}$, $\omega_a(e_0, \ldots, e_{m+n-1})$ is a polynomial in the coefficients of $\alpha^{(i)}$ and $B^{(i)}$ (it is of course also a polynomial in the coefficients of $A^{(i)}$, but we do not need to vary these, and hence for simplicity we shall keep them fixed). By Corollary C.18 the maps $\varphi \mapsto (\alpha^{(0)}, \ldots, \alpha^{(q-1)})$ and $f \mapsto (A^{(0)}, \ldots, A^{(q-1)}, B^{(0)}, \ldots, B^{(q-1)})$ are submersive. Thus to prove the lemma all we need to do is to show that this polynomial is not identically zero, or in other words that for some (f, φ), the corresponding $\omega_a \neq 0$. As in the proof of Proposition 4.3, we shall do this by constructing an (f, φ) for which $\omega_a = \pm a_0 \wedge a_0 \circ A \wedge \cdots \wedge a_0 \circ A^{m-1}$.

Since $T_{1,(x,y)} f_0^{(q)}$ has distinct eigenvalues, we can choose a basis of $T_x M$ consisting of eigenvectors $v_0, \ldots, v_{m-1}$ of $T_{1,(x,y)} f_0^{(q)}$, with corresponding eigenvalues $\lambda_0, \ldots, \lambda_{m-1}$. Similarly choose a basis of $T_y N$ consisting of eigenvectors of $u_m, \ldots, u_{m+n-1}$ of $T_y g^q$, with corresponding eigenvalues $\lambda_m, \ldots, \lambda_{m+n-1}$. Define $u_j = 0$ for $j = 0, \ldots, m-1$; then for any choice of v_j for $j = m, \ldots, m+n-1$ we get a basis $e_0, \ldots, e_{m+n-1}$ for $T_x M \times T_y N$ defined by $e_j = (v_j, u_j)$. We shall construct $v_m, \ldots, v_{m+n-1}$ below.

Now define $e_j^{(i)}$ for $i = 0, \dots, m+n-1$, by $e_j^{(i)} = T_{(x,y)}(f,g)^i(e_j)$. Then since $T_{(x,y)}(f,g)^i$ is an isomorphism, $e_0^{(i)}, \dots, e_{m+n-1}^{(i)}$ is a basis of $T_{x_i}M \times T_{y_i}N$ for each i. Denote $e_j^{(i)} = (v_j^{(i)}, u_j^{(i)})$, note that $v_j^{(i+q)} = \lambda_j v_j^{(i)}$ for $j = 0, \dots, m-1$, and $u_j^{(i+q)} = \lambda_j u_j^{(i)}$ for all j.

Step 1. Let f be any map in $\mathcal{D}(\mathcal{P}_{f_0,g})$ such that $T_{1,(x_i,y_i)} f_0 = T_{1,(x_i,y_i)} f$. Since $v_0^{(i)}, \dots, v_{m-1}^{(i)}$ is a basis for $T_{x_i}M$, we can expand $T_{2,(x_i,y_i)} f(u_j^{(i)})$ for any for $j = m, \dots, m+n-1$ as

$$T_{2,(x_i,y_i)} f(u_j^{(i)}) = B_{j0}^{(i+1)} v_0^{(i+1)} + \cdots + B_{j,m-1}^{(i+1)} v_{m-1}^{(i+1)}.$$

This defines $B_{jk}^{(i)}$ for $j = m, \dots, m+n-1$ and $k = 0, \dots, m-1$, and for all i. Then define

$$\hat{B}_{jk}^{(0)} = \frac{\lambda_k}{\lambda_j - \lambda_k}(B_{jk}^{(1)} + \cdots + B_{jk}^{(q)}).$$

Note that $\lambda_0, \dots, \lambda_{m-1}$ are eigenvalues of $T_{1,(x,y)} f_0^{(q)}$, and $\lambda_m, \dots, \lambda_{m+n-1}$ are eigenvalues of $T_y g^q$, and hence by our conditions on f_0 and g, all the λ_i are distinct; thus $\lambda_j \neq \lambda_k$ if $j \neq k$. Next define $\hat{B}_{jk}^{(i)}$ for all i by

$$\hat{B}_{jk}^{(i+1)} = \hat{B}_{jk}^{(i)} + B_{jk}^{(i+1)}.$$

Finally construct $v_j^{(i)}$ for any i and any $j = m, \dots, m+n-1$ by

$$v_j^{(i)} = \hat{B}_{j0}^{(i)} v_0^{(i)} + \cdots + \hat{B}_{j,m-1}^{(i)} v_{m-1}^{(i)}. \tag{5.2.1}$$

Note that by definition

$$T_{2,(x_i,y_i)} f(u_j^{(i+q)}) = B_{j0}^{(i+q+1)} v_0^{(i+q+1)} + \cdots + B_{j,m-1}^{(i+q+1)} v_{m-1}^{(i+q+1)},$$

and since $v^{(i+q)}k = \lambda_k v_k^{(i)}$ for $k = 0, \dots, m-1$ we have

$$\lambda_j T_{2,(x_i,y_i)} f(u_j^{(i)}) = B_{j0}^{(i+q+1)} \lambda_0 v_0^{(i+1)} + \cdots + B_{j,m-1}^{(i+q+1)} \lambda_{m-1} v_{m-1}^{(i+1)},$$

and so comparing coefficients

$$\lambda_j B_{jk}^{(i+1)} = \lambda_k B_{jk}^{(i+q+1)}.$$

Next we show by induction that

$$\hat{B}_{jk}^{(i)} = \frac{\lambda_k}{\lambda_j - \lambda_k}(B_{jk}^{(i+1)} + \cdots + B_{jk}^{(i+q)}). \tag{5.2.2}$$

This holds for $i = 0$, by definition, so suppose that it holds for some $i > 0$. Then

$$\begin{aligned}
\hat{B}_{jk}^{(i+1)} &= \hat{B}_{jk}^{(i)} + B_{jk}^{(i+1)} \\
&= \lambda_k(\lambda_j - \lambda_k)^{-1}(B_{jk}^{(i+1)} + \cdots + B_{jk}^{(i+q)}) + B_{jk}^{(i+1)} \\
&= \lambda_k(\lambda_j - \lambda_k)^{-1}(B_{jk}^{i+2} + \cdots + B_{jk}^{i+q} + (1 - (\lambda_j - \lambda_k)\lambda_k^{-1}) B_{jk}^{(i+1)}) \\
&= \lambda_k(\lambda_j - \lambda_k)^{-1}(B_{jk}^{i+2} + \cdots + B_{jk}^{i+q} + \lambda_j \lambda_k^{-1} B_{jk}^{(i+1)}) \\
&= \lambda_k(\lambda_j - \lambda_k)^{-1}(B_{jk}^{i+2} + \cdots + B_{jk}^{i+q} + B_{jk}^{(i+q+1)}),
\end{aligned}$$

where the last step follows from $\lambda_j B_{jk}^{(i+1)} = \lambda_k B_{jk}^{(i+q+1)}$. Thus (5.2.2) holds for all $i \geq 0$. But, by definition,

$$\begin{aligned}\hat{B}_{jk}^{(i+q)} &= \hat{B}_{jk}^{(i)} + B_{jk}^{(i+1)} + \cdots + B_{jk}^{(i+q)} \\ &= \hat{B}_{jk}^{(i)} + (\lambda_j - \lambda_k)\lambda_k^{-1}\hat{B}_{jk}^{(i)} \\ &= \lambda_j \lambda_k^{-1} \hat{B}_{jk}^{(i)}.\end{aligned}$$

In other words $\hat{B}_{jk}^{(i)}$ satisfies the same periodicity condition that $B_{jk}^{(i)}$ does:

$$\lambda_k \hat{B}_{jk}^{(i+q)} = \lambda_j \hat{B}_{jk}^{(i)}.$$

Recall that since $T_{1,(x_i,y_i)} f_0 = T_{1,(x_i,y_i)} f$ and $v_0^{(i)}, \ldots, v_{m-1}^{(i)}$ are eigenvectors of $T_{1,(x_i,y_i)} f_0$, we have $v_k^{(i+q)} = \lambda_k v_k^{(i)}$ for $k = 0, \ldots, m-1$. Thus for $j = m, \ldots, m+n-1$,

$$\begin{aligned}v_j^{(i+q)} &= \hat{B}_{j0}^{(i+q)} v_0^{(i+q)} + \cdots + \hat{B}_{j,m-1}^{(i+q)} v_{m-1}^{(i+q)} \\ &= \lambda_j (\hat{B}_{j0}^{(i)} v_0^{(i)} + \cdots + \hat{B}_{j,m-1}^{(i)} v_{m-1}^{(i)}) \\ &= \lambda_j v_j^{(i)}.\end{aligned}$$

Hence $v_j^{(i+q)} = \lambda_j v_j^{(i)}$ for $j = 0, \ldots, m+n-1$, and $e_j^{(i)} = (v_j^{(i)}, u_j^{(i)})$ is an eigenvector of $T_{(x_i,y_i)}(f,g)^q$ for all $j = 0, \ldots, m+n-1$.

Also since $T_{1,(x_i,y_i)} f$ is linear, and $T_{1,(x_i,y_i)} f(v_k^{(i)}) = v_k^{(i+1)}$ for $k = 0, \ldots, m-1$, we have for $j = m, \ldots, m+n-1$

$$\begin{aligned}T_{1,(x_i,y_i)} f(v_j^{(i)}) &= T_{1,(x_i,y_i)} f(\hat{B}_{j0}^{(i)} v_0^{(i)} + \cdots + \hat{B}_{j,m-1}^{(i)} v_{m-1}^{(i)}) \\ &= \hat{B}_{j0}^{(i)} T_{1,(x_i,y_i)} f(v_0^{(i)}) + \cdots + \hat{B}_{j,m-1}^{(i)} T_{1,(x_i,y_i)} f(v_{m-1}^{(i)}) \\ &= \hat{B}_{j0}^{(i)} v_0^{(i+1)} + \cdots + \hat{B}_{j,m-1}^{(i+1)} v_{m-1}^{(i)}.\end{aligned}$$

Then

$$\begin{aligned}v_j^{(i+1)} &= \hat{B}_{j0}^{(i+1)} v_0^{(i+1)} + \cdots + \hat{B}_{j,m-1}^{(i+1)} v_{m-1}^{(i+1)} \\ &= (\hat{B}_{j0}^{(i)} + B_{j0}^{(i+1)}) v_0^{(i+1)} + \cdots + (\hat{B}_{j,m-1}^{(i)} + B_{j,m-1}^{(i+1)}) v_{m-1}^{(i+1)} \\ &= T_{1,(x_i,y_i)} f(v_j^{(i)}) + T_{2,(x_i,y_i)} f(u_j^{(i)}) \\ &= T_{(x_i,y_i)}(f)(v_j^{(i)}, u_j^{(i)})\end{aligned}$$

and hence

$$\begin{aligned}e_j^{(i+1)} &= (v_j^{(i+1)}, u_j^{(i+1)}) \\ &= T_{(x_i,y_i)}(f,g)(v_j^{(i)}, u_j^{(i)}) \\ &= T_{(x_i,y_i)}(f,g)(e_j^{(i)}),\end{aligned}$$

for $j = m, \ldots, m+n-1$; we of course already had this relation for $j = 0, \ldots, m-1$. To conclude, we have shown that with $v_j^{(i)}$ defined for $j = m, \ldots, m+n-1$ using

(5.2.1), it obeys exactly the same relationships as $v_j^{(i)}$ for $j = 0, \ldots, m-1$. Note that we could have defined $v_j^{(i)}$ by $v_j^{(i)} = (\lambda_j Id - T_{1,(x_i,y_i)} f^{(q)})^{-1} T_{2,(x_i,y_i)} f^{(q)}(u_j^{(i)})$; however, we shall need the coordinate-based expression derived above in what follows.

Step 2. Define $\alpha_j^{(i)} = T_{x_i}\varphi(v_j^{(i)})$ for $i = 0, \ldots, q-1$, $j = 0, \ldots, m-1$. Then we claim that $\omega_a(e_0, \ldots, e_{m+n-1})$ is a polynomial in $\{\alpha_j^{(i)}\colon i = 0, \ldots, q-1, j = 0, \ldots, m-1\}$. Recall that $\omega_a = a_0 \wedge \cdots \wedge a_{m+n-1}$ where $a_i = T_{x_i}\varphi \circ T_{(x,y)} f^{(i)}$. Thus $\omega_a(e_0, \ldots, e_{m+n-1}) = \det \alpha$ where $\alpha_{ij} = T_{x_i}\varphi \circ T_{(x_i,y_i)}\pi_1 \circ T_{(x,y)}(f, g)^i(e_j)$. But $T_{(x,y)}(f, g)^i(e_j) = e_j^{(i)} = (v_j^{(i)}, u_j^{(i)})$, and hence $T_{(x_i,y_i)}\pi_1 \circ T_{(x,y)}(f, g)^i(e_j) = v_j^{(i)}$. Thus

$$\alpha_{ij} = T_{x_i}\varphi(v_j^{(i)}),$$

for $i, j = 0, \ldots, m+n-1$. Clearly $\alpha_{ij} = \alpha_j^{(i)}$ for $i = 0, \ldots, q-1$, $j = 0, \ldots, m-1$, and so it remains to show that for other i and j, α_{ij} is a linear combination of the $\alpha_j^{(i)}$. This follows immediately from the relations

$$v_j^{(i)} = \hat{B}_{j0}^{(i)} v_0^{(i)} + \cdots + \hat{B}_{j,m-1}^{(i)} v_{m-1}^{(i)},$$

and

$$v_j^{(i+q)} = \lambda_j v_j^{(i)},$$

and of course the linearity of $T_{x_i}\varphi$.

Step 3. We construct a (f, φ) such that the corresponding $\det \alpha \neq 0$. By Corollary C.18, the map $\varphi \mapsto (T_{x_i}\varphi(v_j^{(i)})\colon i = 0, \ldots, q-1, j = 0, \ldots, m-1)$ is a submersion, and hence a local surjection. Applying this in a neighbourhood of the map $\varphi = 0$, we see that given $\delta > 0$ sufficiently small we can obtain a φ such that for $i = 0, \ldots, q-1$, $j = 0, \ldots, m-1$,

$$T_{x_i}\varphi(v_j^{(i)}) = \alpha_j^{(i)},$$

where we define $\alpha_j^{(i)}$ for $i = 0, \ldots, q-1$, $j = 0, \ldots, m+n-1$ by

$$\alpha_j^{(i)} = \delta \begin{cases} (\lambda_j)^{i+ir}, & \text{if} \quad i \leq s, \\ (\lambda_j)^{s+ir}, & \text{if} \quad i > s, \end{cases}$$

where $m + n = qr + s$ with $0 < s \leq q$. We extend the definition of $\alpha_j^{(i)}$ to $i = q, \ldots, m+n-1$ by

$$\alpha_j^{(i+q)} = \lambda_j \alpha_j^{(i)},$$

for $j = 0, \ldots, m+n-1$. Note that for $i, j = 0, \ldots, m+n-1$, we can write $\alpha_j^{(i)} = a_j^{(0)} \lambda_j^{\sigma(i)}$ where σ is some permutation of $\{0, \ldots, m+n-1\}$ independent of j, such that $\sigma(0) = 0$. Also observe that since $v_j^{(i+q)} = \lambda_j v_j^{(i)}$, we have $T_{x_i}\varphi(v_j^{(i+q)}) = \lambda_j T_{x_i}\varphi(v_j^{(i)}) = \lambda_j \alpha_j^{(i)} = \alpha_j^{(i+q)}$, and hence by induction

$$T_{x_i}\varphi(v_j^{(i)}) = \alpha_j^{(i)},$$

for all $i = 0, \ldots, m+n-1$, and $j = 0, \ldots, m-1$. We shall show below that with a particular choice of $B_{jk}^{(i)}$ we obtain $T_{x_i}\varphi(v_j^{(i)}) = \varepsilon\alpha_j^{(i)}$ for $j = m, \ldots, m+n-1$ and some $\varepsilon > 0$.

Next, we shall need to construct a $f_1 \in \mathcal{D}(\mathcal{P}_{f_0,g})$ such that $T_{2,(x_i,y_i)} f_1 = 0$ for $i = 0, \ldots, q-1$. We do this as follows: Let (U_i, β_i) be a chart for N centred at y_i, such that $\beta_i(U_i)$ is the unit disc in $\mathbb{R}^n$. By shrinking the U_i if necessary, we may assume that they are disjoint, and that the only periodic orbit of g of period less than d contained in U_i is y_i. Let $\Psi_i \colon N \to [0,1]$ be a $\mathcal{C}^\infty$ function such that Ψ_i is identically 1 in a neighbourhood of y_i, and such that its support is contained in U_i. Observe that if $z \in U_i$ then $(1-\Psi_i(z))\beta_i(z) \in \beta_i(U_i)$. Thus we may define $\zeta_i \colon N \to N$ by

$$\zeta(z) = \begin{cases} \beta^{-1}(1-\Psi_i(z)\beta_i(z)); & \text{if} \quad z \in U_i, \\ z, & \text{otherwise.} \end{cases}$$

Note that ζ is $\mathcal{C}^r$ and $\zeta = Id$ outside of the union of the U_i. Furthermore $\zeta(z) = y_i$ in an open neighbourhood of each y_i. Thus define $f_1 \colon M \times N \to M$ by $f_1(w,z) = f_0(w,\zeta(z))$. Then $f_1(\bullet, z) = f_0(\bullet, \zeta(z))$, and hence $f_1(\bullet, z)$ is a diffeomorphism of M for all $z \in N$, thus $f_1 \in \mathcal{D}^r(M \times N, M)$. Also since $\zeta(y) = y$ for any periodic orbit of g of period less than d, we have $f_1(w,z) = f_0(w,z)$ for all $(w,z) \in \mathcal{P}_{f_0,g}$, and hence $f_1 \in \mathcal{D}(\mathcal{P}_{f_0,g})$. Finally, $f_1(w,z)$ is independent of z for z in the neighbourhood of each y_i and hence $T_{2,(x_i,y_i)} f_1 = 0$ as required.

By Corollary C.18, the map $f \mapsto (B^{(0)}, \ldots, B^{(q-1)})$ is submersive on $\mathcal{D}(\mathcal{P}_{f_0,g})$, and hence surjective in a neighbourhood of f_1. Thus given any sufficiently small $\varepsilon > 0$, we can find f such that for $i = 0, \ldots, q-2$,

$$B_{j_0}^{(i+1)} = \varepsilon\left(\frac{\alpha_j^{(i+1)}}{\alpha_0^{(i+1)}} - \frac{\alpha_j^{(i)}}{\alpha_0^{(i)}}\right),$$

with $\alpha_j^{(i)}$ as above, and

$$B_{j_0}^{(q)} = \varepsilon\left(\frac{\lambda_j}{\lambda_0} - \frac{\alpha_j^{(q-1)}}{\alpha_0^{(q-1)}}\right),$$

with $B_{jk}^{(i+1)} = 0$ for all $i = 0, \ldots, q-1$, $j = m, \ldots, m+n-1$, and $k = 1, \ldots, m-1$. Then for all $j = m, \ldots, m+n-1$,

$$\begin{aligned} B_{j_0}^{(1)} + \cdots + B_{j_0}^{(q)} &= \varepsilon\left(\frac{\lambda_j}{\lambda_0} - \frac{\alpha_j^{(q-1)}}{\alpha_0^{(q-1)}} + \frac{\alpha_j^{(q-1)}}{\alpha_0^{(q-1)}} - \frac{\alpha_j^{(q-2)}}{\alpha_0^{(q-2)}} \cdots + \frac{\alpha_j^{(1)}}{\alpha_0^{(1)}} - \frac{\alpha_j^{(0)}}{\alpha_0^{(0)}}\right) \\ &= \varepsilon\left(\frac{\lambda_j}{\lambda_0} - 1\right), \end{aligned}$$

since $a_j^{(0)} = \delta$ for all $j = 0, \ldots, m+n-1$. Thus by (5.2.2),

$$\begin{aligned} \hat{B}_{j_0}^{(0)} &= \lambda_0(\lambda_j - \lambda_0) - 1(B_{j_0}^{(1)} + \cdots + B_{j_0}^{(q)}) \\ &= \varepsilon\lambda_0(\lambda_j - \lambda_0)^{-1}(\lambda_j - \lambda_0)\lambda_0^{-1} \\ &= \varepsilon, \end{aligned}$$

for all $j = 0, \ldots, m+n-1$. Next, we show by induction that

$$\hat{B}_{j0}^{(i)} = \varepsilon \frac{\alpha_j^{(i)}}{\alpha_0^{(i)}}.$$

This holds for $i = 0$ by the above. Suppose now that it holds for some $i > 0$. Then,

$$\begin{aligned}\hat{B}_{j0}^{(i+1)} &= \hat{B}_{j0}^{(i)} + B_{j0}^{(i+1)} \\ &= \varepsilon \frac{\alpha_j^{(i)}}{\alpha_0^{(i)}} + \varepsilon \left(\frac{\alpha_j^{(i+1)}}{\alpha_0^{(i+1)}} - \frac{\alpha_j^{(i)}}{\alpha_0^{(i)}} \right) \\ &= \varepsilon \frac{\alpha_j^{(i+1)}}{\alpha_0^{(i+1)}},\end{aligned}$$

as required. Since by definition $B_{jk}^{(i+1)} = \cdots = B_{jk}^{(i+q)} = 0$ for $k = 1, \ldots, m-1$ and all i, (5.2.2) also gives $\hat{B}_{jk}^{(i)} = 0$ for $k = 1, \ldots, m-1$.

Finally, we can evaluate $T_{x_i}\varphi(v_j^{(i)})$ for $j = m, \ldots, m+n-1$. Since $v_j^{(i)} = \hat{B}_{j0}^{(i)} v_0^{(i)} + \cdots + \hat{B}_{j,m-1}^{(i)} v_{m-1}^{(i)}$, we have

$$v_j^{(i)} = \varepsilon \frac{\alpha_j^{(i)}}{\alpha_0^{(i)}} v_0^{(i)}$$

for all $j = m, \ldots, m+n-1$, and hence since $T_{x_i}\varphi(v_0^{(i)}) = \alpha_0^{(i)}$, we get

$$T_{x_i}\varphi(v_j^{(i)}) = \varepsilon \frac{\alpha_j^{(i)}}{\alpha_0^{(i)}} T_{x_i}\varphi(v_0^{(i)}) = \varepsilon \alpha_j^{(i)}.$$

Thus, as promised, with our choice of $B_{jk}^{(i)}$ we get

$$T_{x_i}\varphi(v_j^{(i)}) = \varepsilon \alpha_j^{(i)},$$

for all $i = 0, \ldots, m+n-1$ and $j = m, \ldots, m+n-1$. Recall that by definition $T_{x_i}\varphi(v_j^{(i)}) = \alpha_j^{(i)}$ for all $i = 0, \ldots, m+n-1$ and $j = 0, \ldots, m-1$ and thus $T_{x_i}\varphi(v_j^{(i)}) = c_j \alpha_j^{(i)}$ for all $i, j = 0, \ldots, m+n-1$, where $c_j = 1$ for $j = 0, \ldots, m-1$ and $c_j = \varepsilon$ for $j = m, \ldots, m+n-1$. Also recall that $\alpha_j^{(i)} = \alpha_j^{(0)} \lambda_j^{\sigma(i)}$ where σ is a permutation of $\{0, \ldots, m+n-1\}$. Since $a_i(e_j) = T_{x_i}\varphi(v_j^{(i)})$ and $\sigma^{(0)} = 0$, we have $a_i(e_j) = c_j \alpha_j^{(0)} \lambda_j^{\sigma(i)} = a_0(e_j)\lambda_j^{\sigma(i)} = a_0(\lambda_j^{\sigma(i)} e_j)$. But e_j is an eigenvector of $A = T_{(x,y)}(f,g)^q$ with eigenvalue λ_j, and so $a_0(\lambda_j^{\sigma(i)} e_j) = a_0(A^{\sigma(i)} e_j)$. Now, $e_0, \ldots, e_{m+n-1}$ are a basis of $T_x M \times T_y N$, and σ is independent of j, and hence $a_i = a \circ A^{\sigma(i)}$ where $a = a_0 = T_x\varphi \circ T_{(x,y)}\pi_1$. Thus for this choice of (f, φ) the first $m+n$ components $a_0, \ldots, a_{m+n-1}$ of $T_x \Phi_{f,\varphi}$ are a permutation of $a_0, a_0 \circ A, \ldots, a_0 \circ A^{m+n-1}$, and hence, as claimed, $\omega_a = \pm a_0 \wedge a_0 \circ A \wedge \cdots \wedge a_0 \circ A^{m-1}$. Since $a_0(e_j) = c_j \alpha_j^{(0)} \neq 0$ for all j, Lemma 4.10 implies that $\omega_a \neq 0$, as required. □

5.3. Construction of $\tilde{W}_I$

Define the map $\tau\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d(M \times N, M^d)$ by

$$\tau^{(f)} = (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(d-1)}).$$

We show that ev_τ is $\mathcal{C}^d$ and compute its tangent map in Corollary C.5 in Appendix C. If we denote the ith component of τ by τ_i, then this gives

$$T_{f,x,y}ev_{\tau_i}(\eta, 0_x, 0_y) = \sum_{j=1}^{i} T_{(x_j,y_j)} f^{(i-j)}(\eta(x_{j-1}, y_{j-1}), 0).$$

As an immediate consequence, we obtain the following.

Lemma 5.12. *$T_{f,x,y}(ev_\tau)$ is surjective at all $(x, y) \notin \mathcal{P}_{f,g}$.*

Proof. If (x, y) is not periodic of period $q < d$, the points $\{(x_i, y_i)\colon i = 0, \ldots, d-1\}$ are distinct, and hence by Corollary C.12 we can find a $\eta \in T_f\mathcal{D}^r(M \times N, M)$ such that η takes on whatever values we want on the (x_i, y_i). Thus, given any $v_i \in T_{x_i}M$ with $1 \le i \le d-1$, choose η such that $\eta(x_{i-1}, y_{i-1}) = v_i$, $\eta(x_i, y_i) = -T_{(x_i,y_i)} f(v_i, 0)$, and $\eta(x_j, y_j) = 0$ for $j \ne i-1, i$. Then $T_{f,x,y}(ev_\tau)(\eta, 0, 0) = (0, \ldots, v_i, \ldots, 0) \in T_{x_0}M \times \ldots \times T_{x_i}M \ldots \times T_{x_{d-1}}M$. Similarly, given any $v_0 \in T_{x_0}M$, pick η such that $\eta(x_0, y_0) = -T_{(x_0,y_0)} f(v_0, 0)$ and $\eta(x_j, y_j) = 0$ for $j \ne 0$. Then by Corollary C.5, $T_{f,x,y}(ev_\tau)(\eta, v_0, 0) = (v_0, 0, \ldots, 0)$. Thus, $T_{f,x,y}(ev_\tau)$ is surjective, as claimed. □

This gives, as a straightforward corollary, the following.

Proof of Proposition 5.4. Define $\nu\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d(M \times N, (M \times N)^d)$ by

$$\nu(f) = (Id, (f, g), (f, g)^2, \ldots, (f, g)^{d-1}).$$

Let Δ_I be the codimension $(d-\alpha)m$ submanifold of M^d defined by

$$\Delta_I = \{(z_0, z_1, \ldots, z_{d-1}) \in M^d\colon z_i = z_{i'} \text{ if and only if } i \sim_I i'\}$$

and $\mathcal{O}$ the codimension 0 (open) submanifold of $(M \times N)^d$

$$\mathcal{O} = \{(w_0, \ldots, w_{d-1}) \in (M \times N)^d\colon w_i \ne w_{i'} \text{ for any } i \ne i'\}.$$

Then $\tilde{W}_I = (\tau(f), \nu(f))^{-1}(\Delta_I \times \mathcal{O})$. By Corollary C.5, ev_τ is $\mathcal{C}^d$, and a similar argument shows that ev_ν is also $\mathcal{C}^d$. Thus, $ev_{(\tau,\nu)}$ is $\mathcal{C}^d$, and $\dim(M \times N) - \operatorname{codim}(\Delta_I \times \mathcal{O}) = m + n - (d-\alpha)m \le m + n < d$. Since $\mathcal{O}$ has codimension 0, $ev_{(\tau,\nu)}$ is transversal to $\Delta_I \times \mathcal{O}$ if and only if ev_τ is transversal to Δ_I for all (x, y) such that $ev_\nu(f, x, y) \in \mathcal{O}$; in other words, all (x, y) such that (x, y) is not periodic under (f, g) with period less than d. But by Lemma 5.12, $T_{f,x,y}(ev_\tau)$ is surjective at all such points, and hence transversal to any submanifold of M^d. Thus $ev_{(\tau,\nu)}$ is transversal to $\Delta_I \times \mathcal{O}$, and hence by the Parametric Transversality Theorem there is a residual set of $f \in \mathcal{D}$ such that $(\tau(f), \nu(f))$ is transversal to $\Delta_I \times \mathcal{O}$, and hence $\tilde{W}_I$ is a codimension $(d-\alpha)m$ submanifold of $(M \times N)\backslash\mathcal{P}_{f,g}$. □

5.4. Construction of $\tilde{T}_{I,I'}$

Proof of Proposition 5.5. Define

$$\Delta_{I,I'} = \{(u_0, u_1, \ldots, u_{d-1}) \in T_{\Delta_I} M^d\colon \text{ if } i \in J_I, \text{ then } u_i \in 0_M \text{ if and only if } i \in I'\},$$

where 0_M is the zero section in TM. Then $\Delta_{I,I'}$ is a codimension $(d-\gamma)m$ submanifold of TM^d. Define the map $\tau'\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^{d-1}(\tilde{T}(M \times N), TM^d)$ by

$$\tau'(f) = (\tilde{T} f^{(0)}, \tilde{T} f^{(1)}, \tilde{T} f^{(2)}, \ldots, \tilde{T} f^{(d-1)}),$$

where as usual $\tilde{T} f^{(i)}$ is the restriction of $Tf^{(i)}$ to $\tilde{T}M$. Let $\nu\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d(M \times N, (M \times N)^d)$ and $\mathcal{O}$ be as in Section 5.3 above. Define $\nu'\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d(\tilde{T}(M \times N), (M \times N)^d)$ by $\nu'(f) = \nu(f) \circ \tau_{M \times N}$, where $\tau_{M\times N}$ is the tangent bundle projection. Thus

$$\nu'(f)(v) = ((x, y), (x_1, y_1), \ldots, (x_{d-1}, y_{d-1})),$$

for all $v \in T_{x,y}(M \times N)$. Then $\tilde{T}_{I,I'} = [\tau'(f), \nu'(f)]^{-1}(\Delta_{I,I'} \times \mathcal{O})$. Lemma C.9 in Appendix C shows that ev'_τ is $\mathcal{C}^{d-1}$ and hence so is $ev_{(\tau',\nu')}$. Also $\dim \tilde{T}(M \times N) - \operatorname{codim}(\Delta_{I,I'} \times \mathcal{O}) = 2m + 2n - 1 - (d-\gamma)m < d - 1$. Thus to prove the proposition it is sufficient to show that $ev_{(\tau',\nu')}$ is transversal to $\Delta_{I,I'} \times \mathcal{O}$ and then apply the Parametric Transversality Theorem. As in Section 5.3, since $\mathcal{O}$ has codimension 0, $ev_{(\tau',\nu')}$ is transversal to $\Delta_{I,I'} \times \mathcal{O}$ if and only if ev'_τ is transversal to $\Delta_{I,I'}$ for all (x, y) such that (x, y) is not periodic under (f, g) with period less than d. In fact, the next lemma shows that $ev_{\tau'}$ is submersive at all such points, and hence is transversal to any submanifold, as required.

Lemma 5.13. *The map $T_{f,v}(ev_{\tau'})$ is surjective at all $v \in \tilde{T}_{x,y}(M \times N)$ with $(x, y) \notin \mathcal{P}_{f,g}$.*

Proof. If (x, y) is not periodic with period less than d, then the points $\{(x_i, y_i)\colon i = 0, \ldots, d-1$ are distinct. Since $(f^{(i)}, g^i)$ is a diffeomorphism and $v \neq 0$, we have $v_i \neq 0$ for all i. Thus by Corollary C.16 we can find a η such that $\omega(T_{(x_i,y_i)}\eta(v_i))$ for $i = 0, \ldots, d-1$ takes on whatever values we want. Hence given some $w_i \in T_{v_i}(TM)$ for $1 \leq i \leq d-1$, choose $\eta \in T_f\mathcal{D}^r(M \times N, M)$ such that $T_{(x_{i-1},y_{i-1})}\eta(v_{i-1}) = \omega(w_i)$, $T_{(x_i,y_i)}\eta(v_i) = -T_{(\eta_i(x,y),0)}(Tf)(\omega(w_i), 0)$ and $T_{(x_j,y_j)}\eta(v_j) = 0$ for $j \neq i-1, i$. Then $T_{f,v}(ev_{\tau'})(\eta, 0) = (0, \ldots, w_i, \ldots, 0) \in T_{v_0}(TM) \times \cdots \times T_{v_i}(TM) \cdots \times T_{v_{d-1}}(TM)$. Also, given any $w_0 \in T_{v_0}(TM)$, take η such that $T_{(x_0,y0)}\eta(v_0) = -T_{\eta_0(x,y),0)}(Tf)(\omega(w_0), 0)$ and $T_{(x_j,y_j)}\eta(v_j) = 0$ for $j \neq 0$. Then $T_{f,v}(ev_{\tau'})(\eta, w_0) = (w_0, \ldots, 0)$. Thus, $T_{f,v}(ev_{\tau'})$ is surjective, as required. □

5.5. *Transversality of* ev_ρ

Proof of Proposition 5.6. Exactly as in Corollary C.8 we have by Lemma B.11

$$T_{\varphi,v}(ev_\rho)(\xi, 0_v) = \begin{pmatrix} \omega(T_{x_{j'_1}}\xi(v_{j'_1})) \\ \omega(T_{x_{j'_2}}\xi(v_{j'_2})) \\ \vdots \\ \omega(T_{x_{j'_{\gamma'}}}\xi(v_{j'_{\gamma'}})) \end{pmatrix},$$

where $v_{j'_i} = T_{(x,y)} f^{(j'_i)}(v)$, and $x_{j'_i} = f^{(j'_i)}(x, y)$. By the definition of $T_{I,I'}$, the points $x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}$ are all distinct, and $v_{j'_i} \neq 0$ for $i = 1, \ldots, \gamma'$. Thus $T_{\varphi,v}(ev_\rho)$ is surjective by Corollary C.16, and hence in particular it is transversal to $L_{I,I'}$ as required. □

5.6. *Construction of* $\tilde{W}_{I,R}$

Proof of Proposition 5.7. Let $\tau\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d(M \times N, M^d)$ and $\nu\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d(M \times N, (M \times N)^d)$ be as in Secton 5.3, and define $\tilde{\tau}\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d((M \times N) \times (M \times N), M^d \times M^d)$ and $\tilde{\nu}\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d((M \times N) \times (M \times N), (M \times N)^d \times (M \times N)^d)$ by $\tilde{\tau}(f) = \tau(f) \times \tau(f)$ and $\tilde{\nu}(f) = \nu(f) \times \nu(f)$, respectively. Thus

$$\begin{aligned}
\tilde{\tau}(f) &= (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(d-1)}) \times (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(d-1)}), \\
\tilde{\nu}(f) &= (Id, (f, g), (f, g)^2, \ldots, (f, g)^{d-1}) \times (Id, (f, g), (f, g)^2, \ldots, (f, g)^{d-1}),
\end{aligned}$$

so that

$$\begin{aligned}
&\tilde{\nu}(f)(x,y,x',y') \\
&\quad = (((x,y), (x_1,y_1), \ldots, (x_{d-1},y_{d-1})), ((x',y'), (x'_1,y'_1), \ldots, (x'_{d-1},y'_{d-1}))).
\end{aligned}$$

Let

$$\begin{aligned}
\Delta_{I,R} = \{&(z_0, \ldots, z_{d-1}, z'_0, \ldots, z'_{d-1}) \in M^d \times M^d\colon (z_0, \ldots, z_{d-1}) \in \Delta_I, \text{ and if} \\
&(i, i') \in J_I \times J_I \text{ then } z_i = z'_{i'} \text{ if and only if } (i, i') \in R\},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{O}' = \{&(w_0, \ldots, w_{d-1}, w'_0, \ldots, w'_{d-1}) \in (M \times N)^d \times (M \times N)^d\colon \\
&w_i \neq w_{i'}, w'_i \neq w'_{i'} \text{ for any } i \neq i' \text{ and } w_i \neq w'_{i'} \text{ for any } i, i'\}.
\end{aligned}$$

Thus $\Delta_{I,R}$ is a codimension $(d-\alpha_R)m$ submanifold of $M^d \times M^d$, and $\mathcal{O}'$ is a codimension 0 (open) submanifold of $(M \times N)^d \times (M \times N)^d$.

Then $\tilde{W}_{I,R} = [\tilde{\tau}(f), \tilde{\nu}(f)]^{-1}(\Delta_{I,R} \times \mathcal{O}')$. By Corollary C.5, $ev_{\tilde{\tau}}$ is $\mathcal{C}^d$, and a similar argument shows that $ev_{\tilde{\nu}}$ is also $\mathcal{C}^d$. Also $\dim(M \times N) \times (M \times N) - \operatorname{codim}(\Delta_I, R \times \mathcal{O}') = 2m+2n-(d-\alpha_R)m \leq 2m+2n < d$. Thus if $ev_{(\tilde{\tau},\tilde{\nu})}$ is transversal to $\Delta_{I,R} \times \mathcal{O}'$, the proposition will follow immediately by the Parametric Transversality Theorem. As in Section 5.3, since $\mathcal{O}'$ has codimension 0, the transversality of $ev_{\tilde{\nu}}$ to $\Delta_{I,R} \times \mathcal{O}'$ is equivalent to the transversality of $ev_{\tilde{\nu}}$ to $\Delta_{I,R}$ at all $\Xi = (f, x, y, x', y')$ such that $\tilde{\nu}(f)(x, y, x', y') \in \mathcal{O}'$. This follows by a slight extension of Lemma 5.12. Thus, $\tilde{\nu}(f)(x, y, x', y') \in \mathcal{O}'$ implies that the points $\{(x_i, y_i)\colon i = 0, \ldots, d-1\} \cup \{(x_i', y_i')\colon i = 0, \ldots, d-1\}$ are disjoint. Hence as in Lemma 5.12, given any $v_i \in T_{x_i}M$, we can find a $\eta \in T f \mathcal{D}^r(M \times N, M)$ such that $T_\Xi(ev_{\tilde{\tau}})(\eta, 0, 0, 0, 0) = ((0_{x_0}, \ldots, v_i, \ldots, 0_{x_{d-1}}), (0_{x_0'}, \ldots, 0_{x_{d-1}'}))$ if $i > 0$ and $T_\Xi(ev_{\tilde{\tau}})(\eta, v_0, 0, 0, 0) = ((v_0, 0_{x_1}, \ldots, 0_{x_{d-1}}), (0_{x_0'}, \ldots, 0_{x_{d-1}'}))$ if $i = 0$. Similarly, given any $v_i' \in T_{x_i'}M$, we can find $\eta' \in T_f\mathcal{D}^r(M \times N, M)$ such that $T_\Xi(ev_{\tilde{\tau}})(\eta', 0, 0, 0, 0) = ((0_{x_0}, \ldots, 0_{x_{d-1}}), (0_{x'0}, \ldots, v_i', \ldots, 0_{x_{d-1}'}))$, if $i > 0$ and $T_\Xi(ev_{\tilde{\tau}})(\eta', 0, 0, v_0', 0) = ((0_{x_1}, \ldots, 0_{x_{d-1}}), (v_0', 0_{x_0'}, \ldots, 0_{x_{d-1}'}))$ if $i = 0$. The set of all such vectors spans $TM^d \times TM^d$, and hence $T_\Xi(ev_{\tilde{\tau}})$ is surjective, as required. □

5.7. Construction of $\tilde{W}_{I,R',(q)}$

Proof of Proposition 5.8. This is a simple modification of Proposition 5.7, using also a part of the proof of Proposition 5.1. First, as in Proposition 5.1, if (x', y') has period q under (f, g) then y' must have period q under g (though not necessarily minimal period q). By our hypothesis on g, such periodic orbits are isolated, so fix one such y'.

Instead of $\tilde{\tau}$ and $\tilde{\nu}$, we shall use $\tilde{\tau}'\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d((M \times N) \times (M \times \{y'\}), M^d \times M^q)$ and $\tilde{\nu}'\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^d((M \times N) \times (M \times \{y'\}), (M \times N)^d \times (M \times N)^q \times M)$ given by

$$\begin{aligned} \tilde{\tau}'(f) &= (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(d-1)}) \times (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(q-1)}), \\ \tilde{\nu}'(f) &= (Id, (f, g), (f, g)^2, \ldots, (f, g)^{d-1}) \\ &\quad \times (Id, (f, g), (f, g)^2, \ldots, (f, g)^{q-1}) \times f^{(q)}. \end{aligned}$$

The additional term $f^{(q)}$ in $\tilde{\nu}'$ will be used to ensure the periodicity of (x', y'). Let

$$\begin{aligned} \Delta_{I,R'} = \{&(z_0, \ldots, z_{d-1}, z_0', \ldots, z_{q-1}') \in M^d \times M^q\colon (z_0, \ldots, z_{d-1}) \in \Delta_I, \text{ and} \\ &\text{if } (i, i') \in J_I \times \{0, \ldots, q-1\} \text{ then } z_i = z_{i'}' \text{ if and only if } (i, i') \in R'\}, \end{aligned}$$

and

$$\begin{aligned} \mathcal{O}q = \{&(w_0, \ldots, w_{d-1}, w_0', \ldots, w_{q-1}', z_q') \in (M \times N)^d \times (M \times N)^q \times M\colon \\ &w_i \neq w_{i'} \text{ for any } i \neq i', w_i \neq w_{i'}' \text{ for any } i, i', \\ &w_i' \neq w_{i'}' \text{ for any } i \neq i' \text{ such that } 0 \leq i, i' < q, \text{ and} \\ &\pi_1(w_0') = z_q'\}, \end{aligned}$$

where $\pi_1\colon M \times N \to M$ is projection onto the first factor. Thus $\Delta_{I,R'}$ is submanifold of $M^d \times M^q$ of dimension $(\alpha + q - \beta_{R'})m$, and hence of codimension $(d + q)m - (\alpha + q - \beta_{R'})m = (d - \alpha_{R'})m$, whilst $\mathcal{O}_q$ is a codimension m submanifold of $(M \times N)^d \times (M \times N)^q \times M$. Hence $\Delta_{I,R'} \times \mathcal{O}_q$ has codimension $(d + 1 - \alpha_{R'})m$.

Note that if $ev_{\tilde{\nu}'}(f, x, y, x', y') \in \mathcal{O}_q$, then $(f, g)^q(x', y') = (f^{(q)}(x', y'), g^q(y')) = (x', y')$, but $(f, g)^i(x', y') \neq (x', y')$ for all $i = 1, \ldots, q-1$, and hence (x', y') is periodic of minimal period q. Thus $\tilde{W}_{I,R',(q)} = [\tilde{\tau}'(f), \tilde{\nu}'(f)]^{-1}(\Delta_{I,R'} \times \mathcal{O}_q)$. Hence if $(\tilde{\tau}'(f), \tilde{\nu}'(f))$ is transversal to $\Delta_{I,R'} \times \mathcal{O}_q$, then $\tilde{W}_{I,R',(q)}$ will be a codimension

$(d+1-\alpha_{R'})m$ submanifold of $(M\times N)\times(M\times\{y'\})$, and hence a codimension $(d-\alpha_{R'})m+m+n$ submanifold of $(M\times N)\times(M\times N)$.

As in Section 5.6, $ev_{\tilde{\tau}'}$ and $ev_{\tilde{\nu}'}$ are $\mathcal{C}^d$. Also $\dim(M\times N)\times(M\times N)-\operatorname{codim}(\Delta_{I,R'}\times\mathcal{O}_q)=2m+2n-(d-\alpha_{R'})m\le 2m+2n<d$. Thus if $ev_{(\tilde{\tau}',\tilde{\nu}')}$ is transversal to $\Delta_{I,R'}\times\mathcal{O}_q$, the proposition will follow immediately by the Parametric Transversality Theorem. Unfortunately, since $ev_{\tilde{\nu}'}$ is not automatically transversal to $\mathcal{O}_q$, it is no longer sufficient to concentrate just on $ev_{\tilde{\tau}'}$, as it was in Section 5.6. We thus proceed as follows.

First, observe that if $ev_{\tilde{\nu}'}(f,x,y,x',y')\in\mathcal{O}_q$, then the points $\{(x_i,y_i)\colon i=0,\dots,d-1\}\cup\{(x_i',y'i)\colon i=0,\dots,q-1\}$ are all disjoint. Thus, as in the proof of Proposition 5.7, given any $v_i\in T_{x_i}M$, $i>0$, we can find a $\eta\in T_f\mathcal{D}^r(M\times N,M)$ such that $\eta(x_{i-1},y_{i-1})=v_i$, $\eta(x_i,y_i)=-T_{(x_i,y_i)}f(v_i,0)$, $\eta(x_j,y_j)=0$ for $j\ne i-1,i$ and $\eta(x_j',y_j')=0$ for $j=0,\dots,q-1$. Then if we denote $\Xi=(f,x,y,x',y')$, we have $T_\Xi(ev_{\tilde{\tau}'})(\eta,0,0,0,0)=((0_{x_0},\dots,v_i,\dots,0_{x_{d-1}}),0_{z'})$ where $0_{z'}=(0_{x_0'},\dots,0_{x_{q-1}'})$. Furthermore, since $(f,g)^i=(f^{(i)},g^q)$, we have $T_\Xi(ev_{\tilde{\nu}'})(\eta,0,0,0,0)=((0_{x_0},0_{y_0}),\dots,(v_i,0_{y_i}),\dots,(0_{x_{d-1}},0_{y_{d-1}}),(0_{w'},0_{x_q'}))$ where $0_{w'}=((0_{x_0'},0_{y_0'}),\dots,(0_{x_{q-1}'},0_{y_{q-1}'}))$. Similarly, given any $v_0\in T_{x_0}M$, pick η such that $\eta(x_0,y_0)=-T_{(x_0,y_0)}f(v_0,0)$, $\eta(x_j,y_j)=0$ for $j\ne 0$ and $\eta(x_j',y_j')=0$ for $j=0,\dots,q-1$. Then $T_\Xi(ev_{\tilde{\tau}'})(\eta,v_0,0,0,0)=((v_0,\dots,0_{x_{d-1}}),0_{z'})$ and $T_\Xi(ev_{\tilde{\nu}'})(\eta,v_0,0,0,0)=((v_0,0_{y_0}),\dots,(0_{x_{d-1}},0_{y_{d-1}}),(0_{w'},0_{x_q'}))$. Now, $((u_0,\dots,u_{d-1}),0_{w'},0_{x_q'})\in T\mathcal{O}_q$ for any $(u_0,\dots,u_{d-1})$ with $u_i\in T_{w_i}(M\times N)$ and hence we see that $\operatorname{Image}(T_\Xi(ev_{\tilde{\tau}',\tilde{\nu}'}))+T(\Delta_{I,R'}\times\mathcal{O}_q)$ contains $TM^d\times\{0_{z'}\}\times T(M\times N)^d\times\{0_{w'}\}\times\{0_{x_q'}\}$.

Similarly, given $v_i'\in T_{x_i'}M$, we can find a $\eta\in T_f\mathcal{D}^r(M\times N,M)$ such that $T_\Xi(ev_{\tilde{\tau}'})(\eta,0,0,0,0)=(0_z,(0_{x_0'},\dots,v_i',\dots,0_{x_{q-1}'}))$ if $i>0$, and $T_\Xi(ev_{\tilde{\tau}'})(\eta,0,0,v_0',0)=(0_z,(v_0',0_{x_0'},\dots,0_{x_{q-1}'}))$ if $i=0$, where $0_z=(0_{x_0},\dots,0_{x_{d-1}})$. As before we have $T_\Xi(ev_{\tilde{\nu}'})(\eta,0,0,0,0)=(0_w,((0_{x_0'},0_{y_0'}),\dots,(v_i',0_{y_i}),\dots,(0_{x_{q-1}'},0_{y_{q-1}'})),0_{x_q'})$ if $i>0$, and $T_\Xi(ev_{\tilde{\nu}'})(\eta,0,0,v_0',0)=(0_w,((v_0',0_{y_0'}),\dots,(0_{x_{q-1}'},0_{y_{q-1}'})),0_{x_q'})$ if $i=0$, where $0_w=((0_{x_0},0_{y_0}),\dots,(0_{x_{d-1}},0_{y_{d-1}}))$.

However, at this stage we cannot proceed as above, since it is not true that $(0_w,(u_0',\dots,u_{q-1}'),0_{x_0'})\in T\mathcal{O}_q$ for any $(u_0',\dots,u_{q-1}')$ with $u_i'\in T_{w_i'}(M\times N)$. Instead we have $(0_w,(u_0',\dots,u_{q-1}'),\pi_1(u_0'))\in T\mathcal{O}_q$. Fortunately, by the proof of Proposition 5.1, if $f\in\mathcal{D}^r$ and (x',y') is periodic of period q, then given any $v_q'\in T_{x_q'}M$, we can find $\eta\in T_f\mathcal{D}^r(M\times N,M)$ such that $T_{f,x',y'}(ev_\rho)(\eta,0)=(v_q',0)$ where $\rho(f)=(f^{(q)},f^{(0)})$. Furthermore, since the set $\{(x_i,y_i)\colon i=0,\dots,d-1\}$ is disjoint from $\{(x_i',y_i')\colon i=0,\dots,q\}$, we can choose $\eta(x_i,y_i)=0$ for all i. For such a η, we then have $T_\Xi(ev_{\tilde{\nu}'})(\eta,0,0,0,0)=(0_w,((0_{x_0'},0_{y_0'}),\tilde{u}_1,\dots,\tilde{u}_{q-1}),v_q')$ for some $\{\tilde{u}_1',\dots,\tilde{u}_{q-1}'\}$ (depending on v_q'). So now, given any $((u_0',\dots,u_{q-1}'),v_q')$, choose $\eta\in T_f\mathcal{D}^r(M\times N,M)$ such that $T_\Xi(ev_{\tilde{\nu}'})(\eta,0,0,0,0)=(0_w,((0_{x_0'},0_{y_0'}),\tilde{u}_1',\dots,\tilde{u}_{q-1}'),v_q'-\pi_1(u_0'))$ for some $\{\tilde{u}_1',\dots,\tilde{u}_{q-1}'\}$. Since we have $(0_w,(u_0',u_1'-\tilde{u}_1',\dots,u_{q-1}'-\tilde{u}_{q-1}'),\pi_1(u_0'))\in T\mathcal{O}_q$. Adding these together, we conclude that $(0_w,((u_0',\dots,u_{q-1}'),v_q'))\in\operatorname{Image}(T_\Xi(ev_{\tilde{\nu}'}))+T\mathcal{O}_q$ as required.

We thus conclude that $\operatorname{Image}(T_\Xi(ev_{(\tilde{\tau}',\tilde{\nu}')})+T(\Delta_{I,R'}\times\mathcal{O}_q)$ contains $\{0_z\}\times TM^d\times\{0_w\}\times T(M\times N)^d\times TM$. Since we have already shown that it contains $TM^d\times\{0_{z'}\}\times T(M\times N)^d\times\{0_{w'}\}\times\{0_{x_q'}\}$, we conclude that $ev_{(\tilde{\tau},\tilde{\nu}')}$ is transversal to $\Delta_{I,R'}\times\mathcal{O}_q$, as required. ☐

5.8. *Transversality of* ev_σ

Proof of Proposition 5.9. As in Corollary C.3, we have

$$T_\Xi(ev_\sigma)(x,0,0,0,0) = \left(\begin{pmatrix} \xi(x_{j'_1}) \\ \xi(x_{j'_2}) \\ \vdots \\ \xi(x_{j'_{\gamma'}}) \end{pmatrix}, \begin{pmatrix} \xi(x'_{j'_1}) \\ \xi(x'_{j'_2}) \\ \vdots \\ \xi(x'_{j'_{\gamma'}}) \end{pmatrix} \right).$$

By the definition of $\tilde{W}_{I,R}$ the points $x_{j'_1}, x_{j'_2}, \ldots, x_{j'_\gamma}$ are distinct and $\{x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}\} \cap \{x'_{j'_1}, x'_{j'_2}, \ldots, x'_{j'_{\gamma'}}\} = \varnothing$. Thus, using Corollary C.12, given any $i \in J_{I,R}$, we can find a $\xi_i \in T_\varphi(\mathcal{C}^r(M,\mathbb{R}))$ such that $T_\Xi(ev_\sigma)(\xi,0,0,0,0) = (e_i,0)$, where e_i is the ith unit vector in $T\mathbb{R}^{\gamma''}$. Also $(e_i,e_i) \in T\hat{\Delta}_{\gamma''}$ for any i, and hence Image$(T_\Xi(ev_\sigma)) + T\hat{\Delta}_{\gamma''} = T\mathbb{R}^{\gamma''} \times T\mathbb{R}^{\gamma''}$, as required. □

5.9. *Transversality of* $ev_{\sigma'}$

Proof of Proposition 5.10. As in Section 5.8,

$$T_\Xi(ev_{\sigma'})(\xi,0,0,0,0) = \left(\begin{pmatrix} \xi(x_{j'_1}) \\ \xi(x_{j'_2}) \\ \vdots \\ \xi(x_{j'_{\gamma'}}) \end{pmatrix}, \begin{pmatrix} \xi(x'_{j'_1}) \\ \xi(x'_{j'_2}) \\ \vdots \\ \xi(x'_{j'_{\gamma'}}) \end{pmatrix} \right).$$

If $(x,y,x',y') \in \tilde{W}_{I,R',(q)}$ are distinct and $\{x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}\} \cap \{x'_0, x'_1, \ldots, x'_{q-1}\} = \varnothing$. But (x',y') is periodic with period q, and hence $x'_j \in \{x'_0, x'_1, \ldots, x'_{q-1}\}$ for any $j \in \{j'_1, j'_2, \ldots, j'_{\gamma'}\}$. Hence $\{x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}\} \cap \{x'_{j'_1}, x'_{j'_2}, \ldots, x'_{j'_{\gamma'}}\} = \varnothing$. Thus, as in Section 5.8, given any $i \in J_{I,R'}$, we can find a $\xi_i \in T_\varphi(\mathcal{C}^r(M,\mathbb{R}))$ such that $T_\Xi(ev_{\sigma'})(\xi,0,0,0,0) = (e_i,0)$, and thus Image$(T_\Xi(ev_{\sigma'})) + T\hat{\Delta}_{\gamma''} = T\mathbb{R}^{\gamma''} \times T\mathbb{R}^{\gamma''}$ as required. □

5.10. *Transversality of* $ev_{\sigma''}$

Proof of Proposition 5.11. As usual, we have

$$T_\Xi(ev_{\sigma''})(\xi,0,0,0,0) = \left(\begin{pmatrix} \xi(x_{j'_1}) \\ \xi(x_{j'_2}) \\ \vdots \\ \xi(x_{j'_{\gamma'}}) \end{pmatrix}, \begin{pmatrix} \xi(x'_{j'_1}) \\ \xi(x'_{j'_2}) \\ \vdots \\ \xi(x'_{j'_{\gamma'}}) \end{pmatrix} \right),$$

with $x'_j = x_{j+k}$. If $(x,y,x',y') \in \tilde{W}_{\tilde{I},(k)}$, then the points $\{x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\tilde{\gamma}}}\}$ are distinct. By Corollary C.12, given any $j'_i \in J_{\tilde{I}}$, we can find a $\xi_j \in T_\varphi(\mathcal{C}^r(M,\mathbb{R}))$ such that

$\xi_{j'_i}(x_{j'_i}) = 1$ and $\xi_{j'_i}(x_j) = 0$ for all $j \in \{0, \ldots, d+k-1\}$ such that $x_{j'_i} \neq x_j$, in other words, such that $j \notin \tilde{I}_i$. In particular $x_{j'_i} \neq x_{j'_{i'}}$ for all $j'_{i'} \in J_{\tilde{I}}$ such that $i \neq i'$. Thus $T_\Xi(ev_{\sigma''})(\xi_{j'_i}, 0, 0, 0, 0) = (e_i, u^{(i)})$, where e_i is the ith unit vector in $T\mathbb{R}^{\tilde{\gamma}}$ and the (i')th component of $u^{(i)}$ is 1 if and only if $x_{j'_i} = x'_{j'_{i'}} = x\, j'_{i'+k}$. But $x_{j'_i} = x_{j'_{i'+k}}$ if and only if $j_i'' + k \in \tilde{I}_i$, and thus

$$u^{(i)} = \sum_{j'_{i'}:\ j'_{i'}+k \in \tilde{I}_i} e_{i'}.$$

Now, recall that j'_i is the maximal element of $\tilde{I}_i$ and $k > 0$. Thus if $j'_{i'} + k \in \tilde{I}_i$, then $j'_{i'} < j'_i$ and hence $i' < i$. In particular $u^{(1)} = 0$, and hence $T_\Xi(ev_{\sigma''})(\xi_1, 0, 0, 0, 0) = (e_1, 0)$. Now proceed by induction. Our inductive hypothesis is that for a given i with $1 < i < \tilde{\gamma}$, we have $(e_{i'}, 0) \in T\hat{\Delta}_{\tilde{\gamma}} + \text{Image}(T_\Xi(ev_{\sigma''}))$ for all i' such that $i' < i$. Then by the above, $(e_i, u^{(i)}) \in \text{Image}(T_\Xi(ev_{\sigma''}))$. Since $j'_{i'} + k \in \tilde{I}$ implies $i' < i$, we see that $u^{(i)}$ is the sum of a set of $e_{i'}$ over some set of $i' < i$. Hence the inductive hypothesis implies that $(u^{(i)}, 0) \in T\hat{\Delta}_{\tilde{\gamma}} + \text{Image}(T_\Xi(ev_{\sigma''}))$. But $(u^{(i)}, u^{(i)}) \in T\hat{\Delta}_{\tilde{\gamma}}$. Hence $(e_i, 0) = (e_i, u^{(i)}) + (u^{(i)}, 0) - (u^{(i)}, u^{(i)}) \in T\hat{\Delta}_{\tilde{\gamma}} + \text{Image}(T_\Xi(ev_{\sigma''}))$, as required.

Thus by induction, $(e_i, 0) \in T\hat{\Delta}_{\tilde{\gamma}} + \text{Image}(T_\Xi(ev_{\sigma''}))$ for all i, and since $(e_i, e_i) \in T\hat{\Delta}_{\tilde{\gamma}}$, we conclude that $T\hat{\Delta}_{\tilde{\gamma}} + \text{Image}(T_\Xi(ev_{\sigma''})) = T\mathbb{R}^{\tilde{\gamma}} \times \mathbb{R}^{\tilde{\gamma}}$, and hence $ev_{\sigma''}$ is transversal to $T\hat{\Delta}_{\tilde{\gamma}}$, as required. □

5.11. Periodically Forced Differential Equations

The proof of Theorem 3.3 has exactly the same structure as the proof of Theorem 3.1, replacing the maps $f^{(i)}$ by $f^{\tau i}$. Essentially the only modification required is that instead of showing that ev_τ, $ev_{\tau'}$ are surjective on appropriate domains (as in Sections 5.3 and 5.4), we need to demonstrate the surjectivity of the maps $(\Psi, x, \theta) \mapsto (x, f^\tau(x, \theta), \ldots, f^{\tau(d-1)}(x, \theta))$ and $(\Psi, v) \mapsto (x, Tf^\tau(v), \ldots, Tf^{\tau(d-1)}(v))$. This follows by straightforward extensions of Lemmas 31.7 and 31.11 of Abraham and Robbin [1967]. Similar modifications are also required to Sections 5.6 and 5.7.

6. Proof of the Bundle Embedding Theorem

The proof of Theorem 3.2 is closely related to the proof of the Forced Takens Theorem (Theorem 3.1) given in Section 5. Thus, the first part of the argument is to show that for a residual set of f and φ, $\tilde{T}\Phi_{f,g,\varphi}$ is transversal to the zero section in $T\mathbb{R}^d$ and $\Phi_{f,g,\varphi} \times \Phi_{f,g,\varphi}$ is transversal to the diagonal in $\mathbb{R}^d \times \mathbb{R}^d$. The basic idea behind this is the same as that in Section 5; though there are a number of useful simplifications: (a) we can completely ignore periodic orbits; (b) when analysing $\Phi_{f,g,\varphi} \times \Phi_{f,g,\varphi}$ we can restrict to pairs of points (x, y) and (x', y') such that $y = y'$; (c) we do not need to treat pairs such that $(x', y') = (f, g)^k(x, y)$; and (d) when proving immersivity we can restrict to vectors of the form $(v, 0_y)$ and, since $Tf^{(i)}(v, 0_y) \neq 0$ if $v \neq 0$, we need not bother with the construction of the $\tilde{T}_{I,I'}$.

Once we have shown the transversality of $\tilde{T}\Phi_{f,g,\varphi}$ and $\Phi_{f,g,\varphi} \times \Phi_{f,g,\varphi}$, we complete the proof by treating y as a parameter and applying the Parametric Transversality Theorem to the maps $y \mapsto \tilde{T}\Phi_{f,g,\varphi,y}$ and $y \mapsto \Phi_{f,g,\varphi,y} \times \Phi_{f,g,\varphi,y}$. This shows that $\tilde{T}\Phi_{f,g,\varphi,y}$ and $\Phi_{f,g,\varphi,y} \times \Phi_{f,g,\varphi,y}$ are transversal to the zero section in $T\mathbb{R}^d$ and the diagonal in $\mathbb{R}^d \times \mathbb{R}^d$, respectively, for a residual set of y. By counting dimensions, this immediately implies the immersivity and injectivity of $\Phi_{f,g,\varphi,y}$, as required.

Exactly as in Section 5, the main obstacle is generated by points such that $x_i = x_j$ with $i \neq j$. Unfortunately, we cannot use exactly the same approach as we did there to deal with these points. To explain this, consider just the case of immersivity. Suppose we try and proceed as before, so that we define submanifolds $\tilde{W}_I$ each characterized by the set of pairs (i, j) for which $x_i = x_j$. We can certainly then show that $T\Phi_{\varphi,I}$ is transversal to the zero section in $T\mathbb{R}^\alpha$ and that $\tilde{W}_{I,y} = \tilde{W}_I \cap (M \times \{y\})$ is a codimension $(d-\alpha)m$ submanifold for a residual set of y. Thus it seems plausible that generically $\Phi_{\varphi,I}$ should be immersive on $\tilde{W}_{I,y}$ if $\alpha \geq 2(m - (d-\alpha)m)$, which is always satisfied if $d \geq 2m+1$. The problem arises in the definition of the map $y \mapsto \tilde{T}\Phi_{\varphi,I,y}$. The domain of $\tilde{T}\Phi_{\varphi,I,y}$ is $T\tilde{W}_{I,y} = T\tilde{W}_I \cap (TM \times \{y\})$ and hence depends on y. This makes it impossible to apply the Parametric Transversality Theorem directly. A similar problem arises with $y \mapsto \Phi_{\varphi,I,R} \times \Phi_{\varphi,I,R}$. One way around this difficulty would be to extend the Parametric Transversality Theorem to cover this case. We shall use a more direct approach that avoids this issue by combining the construction of the $\tilde{W}_I$ with the proof of transversality. As we have already remarked in the introduction to Section 5, the same method can be used there.

6.1. Preliminaries

As in the previous section, we only need prove the theorem for sufficiently large r; in this case $r \geq 2n$. However, the passage to lower degrees of smoothness is not quite so trivial in this case, and hence is given in a separate section (Section 6.4).

Recall that $\Phi_{f,g,\varphi,y}(x) = \Phi_{f,g,\varphi}(x, y)$; since g is assumed fixed, we shall drop it from the notation and write $\Phi_{f,\varphi,y}(x)$ and $\Phi_{f,\varphi}$, respectively. Since we only need to prove the theorem for a set of y of full Lebesgue measure, we can completely ignore the closure $\bar{\mathcal{P}}_g$ of the set of periodic orbits of g of period less than d. Thus instead of working with $(M \times N)\backslash\mathcal{P}_{f,g}$, we consider $M \times (N\backslash\bar{\mathcal{P}}_g)$. For notational convenience, we shall define $\tilde{N} = N\backslash\bar{\mathcal{P}}_g$. By our hypothesis this is an open dense submanifold of N, of Lebesgue measure 1.

Note that since $\Phi_{f,g,\varphi,y}$ depends continuously on y and embeddings are open in $\mathcal{C}^r(M, \mathbb{R}^d)$, the set of y such that $\Phi_{f,g,\varphi,y}$ is an embedding is open in both $\tilde{N}$ and N. Furthermore a set of Lebesgue measure is necessarily dense (as otherwise its complement would contain an open set and thus have strictly positive measure). Thus it suffices to prove that the set of y such that $\Phi_{f,g,\varphi,y}$ is an embedding has full Lebesgue measure in $\tilde{N}$ (and hence in N) for a residual set of f and φ.

As in Section 5, let $I = \{I_1, I_2, \dots, I_a\}$ be a partition of $\{0, \dots, d-1\}$, and $\sim_I$ be the equivalence relation defined by $i \sim_I i'$ if and only if i, i' are in the same element of the partition. Recall that J_I is a set containing precisely one element from each I_k for $k = 1, \dots, \alpha$. As usual, write these as $J_I = \{j_1, j_2, \dots, j_\alpha\}$ with $j_1 < j_2 < \cdots < j_\alpha$.

6.2. Immersivity of Φ

As in Section 5, the basic idea is to make $\tilde{T}\Phi_{f,\varphi}$ transversal to the zero section in $T\mathbb{R}^d$ and then count dimensions. For a given $y \in \tilde{N}$, define the maps $\Phi_{f,\varphi,I}\colon M \times N \to \mathbb{R}^\alpha$ and $\Phi_{f,\varphi,I,y}\colon M \times \{y\} \to \mathbb{R}^\alpha$ by

$$\begin{aligned}\Phi_{f,\varphi,I}(x,y) &= (\varphi(x_{j_1}), \varphi(x_{j_2}), \ldots, \varphi(x_{j_\alpha}))^\dagger,\\ \Phi_{f,\varphi,I,y}(x) &= \Phi_{f,\varphi,I}(x,y),\end{aligned}$$

where $x_i = f^{(i)}(x,y)$. Observe that if for some $v \in \tilde{T}_x M$ we have $\Phi_{f,\varphi,I,y}(v) \neq 0$, then $T_x\Phi_{f,\varphi,y}(v) \neq 0$. Hence if for all $v \in \tilde{T}_x M$, we have $T_x\Phi_{f,\varphi,I,y}(v) \neq 0$ for some I, then $\Phi_{f,\varphi,y}$ is immersive at x.

Define the map $\rho\colon \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M,\mathbb{R}) \to \mathcal{C}^{n-1}(\tilde{T}M \times \tilde{N}, T\mathbb{R}^\alpha)$ by

$$\begin{aligned}\rho(f,\varphi)(v,y) &= T_x\Phi_{f,\varphi,I,}(v,0y)\\ &= T_x\Phi_{f,\varphi,I,y}(v),\end{aligned}$$

and as usual let L_I be the zero section in $T\mathbb{R}^\alpha$. Since $r \geq 2n$, ev_ρ is $\mathcal{C}^{n-1}$ by Lemma C.9. As in Section 5.3, define $\tau\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^n(M \times N, M^d)$ by $\tau(f) = (f^{(0)}, f^{(1)}, f^{(2)}, \ldots, f^{(d-1)})$. Let Δ_I be the codimension $(d-\alpha)m$ submanifold of M^d defined as before by

$$\Delta_I = \{(z_1, z_2, \ldots, z_d) \in M^d\colon z_i = z_{i'} \text{ if and only if } i \sim_I i'\}.$$

Note that since we replace $(M \times N)\backslash\mathcal{P}_{f,g}$ by $M \times \tilde{N} = M \times (N\backslash\bar{\mathcal{P}}_g)$, which is independent of f, we can dispense with ν and $\mathcal{O}$. Then Corollary C.5 implies that ev_τ is $\mathcal{C}^n$. Define $\tau'\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^n(\tilde{T}M \times \tilde{N}, M^d)$ by $\tau'(f) = \tau(f) \circ \tau_M$ where $\tau_M\colon \tilde{T}M \mapsto M$ is the tangent bundle projection. Then $ev_{\tau'}$ is also $\mathcal{C}^n$. Finally define $\rho'\colon \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M,\mathbb{R}) \to \mathcal{C}^n(\tilde{T}M \times \tilde{N}, T\mathbb{R}^\alpha \times M^d)$ by

$$\begin{aligned}\rho'(f,\varphi)(v,y) &= (\rho(f,\varphi)(v,y), \tau'(f)(v,y))\\ &= (T_x\Phi_{f,\varphi,I,y}(v), \tau(f)(x,y)),\end{aligned}$$

where $x = \tau_M(v)$, i.e., $v \in T_x M$.

Proposition 6.1. *The map $ev_{\rho'}$ is transversal to $L_I \times \Delta_I$ for all I.*

Proof. Suppose that $ev_{\rho'}(f,\varphi,v,y) \in L_I \times \Delta_I$. Then $\rho(f,\varphi)(v,y) \in L_I$, so $\rho(f,\varphi)(v,y) = 0$, and we may write $ev_{\rho'}(f,\varphi,v,y) = (0,z)$ for some $z \in \Delta_I$. Since $\mathcal{P}_{f,g} \subset M \times \mathcal{P}_g$, if $(x,y) \in M \times \tilde{N}$ then $(x,y) \notin \mathcal{P}_{f,g}$. Lemma 5.12 thus implies that $T_{f,v,y}(ev_{\tau'})$ is surjective and hence given $u' \in T_z\Delta_I$, there exists a $(\eta, w) \in T_f\mathcal{D}^r(M \times N, M) \times T(\tilde{T}M)$ such that $T_{f,v,y}(ev_{\tau'})(\eta, w, 0_y) = u'$. Let $\Xi = (f,\varphi,v,y)$. Then for any $\xi \in T_\varphi\mathcal{C}^r(M,\mathbb{R})$, we have by linearity

$$\begin{aligned}T_\Xi(ev_\rho)(\eta, x, w, 0_y) &= T_\Xi(ev_\rho)(\eta, 0_\varphi, w, 0_y) + T_\Xi(ev_\rho)(0_f, \xi, 0_v, 0_y)\\ &= u_0 + T_\Xi(ev_\rho)(0_f, \xi, 0_v, 0_y),\end{aligned}$$

for some $u_0 \in T(T\mathbb{R}^\alpha)$, and just as in Section 5.5,

$$T_\Xi(ev_\rho)(0_f, \xi, 0_v, 0_y) = \begin{pmatrix} \omega(T_{x_{j_1}}\xi(v_{j_1})) \\ \omega(T_{x_{j_2}}\xi(v_{j_2})) \\ \vdots \\ \omega(T_{x_{j_\alpha}}\xi(v_{j_\alpha})) \end{pmatrix},$$

where $x_j = f^{(j)}(x, y)$ as usual but now $v_j = T_{(x,y)} f^{(j)}(v, 0_y)$. Since $\tau(f)(x, y) \in \Delta_I$, the points $x_{j_1}, x_{j_2}, \dots, x_{j_\gamma}$ are all distinct. For a fixed y, $f^{(j_i)}$ is a diffeomorphism and $\|v\| = 1$ and hence $v_{j_i} \neq 0$ for $i = 1, \dots, \gamma$. Thus, by Corollary C.16, given any $u \in T_0 L_I$, there exists a $\xi \in T_\varphi \mathcal{C}^r(M, \mathbb{R})$ such that $T_\Xi(ev_\rho)(0_f, \xi, 0_v, 0_y) = u - u_0$. Hence $T_\Xi(ev_\rho)(\eta, \xi, w, 0_y) = u$ and so $T_\Xi(ev_{\rho'})(\eta, \xi, w, 0_y) = (u, u')$. Thus $T_\Xi(ev_{\rho'})$ is surjective, and in particular transversal to $L_I \times \Delta_I$, as required. □

Observe that $\dim \tilde{T}M \times \tilde{N} - \operatorname{codim} L_I \times \Delta_I = 2m - 1 + n - (d - \alpha)m - a$, and

$$\begin{aligned} 2m - 1 + n - (d-\alpha)m - a &\le d - 2 - (d-\alpha)m - a + n \\ &\le (d-\alpha)(1-m) - 2 + n \\ &< n - 1, \end{aligned} \tag{6.2.1}$$

and $ev_{\rho'}$ is $\mathcal{C}^{n-1}$. Thus by the Parametric Transversality Theorem there is a residual set of $(f, \varphi) \in \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$ for which $\rho'(f, \varphi)$ is transversal to $L_I \times \Delta_I$. Fix any such (f, φ), and define $\rho'' \colon \tilde{N} \to \mathcal{C}^{n-1}(\tilde{T}M, T\mathbb{R}^\alpha \times M^d)$ by

$$\begin{aligned} \rho''(y)(v) &= \rho'(f, \varphi)(v, y) \\ &= (T_x \Phi_{f,\varphi,I,y}(v), \tau(f)(x, y)). \end{aligned}$$

Then $ev_{\rho''}(y, v) = \rho'(f, \varphi)(v, y)$ and hence $ev_{\rho''}$ is transversal to $L_I \times \Delta_I$. Also $ev_{\rho''}$ is $\mathcal{C}^n$, and using (6.2.1), we have $\dim \tilde{T}M - \operatorname{codim} L_I \times \Delta_I = 2m - 1 - (d-\alpha)m - \alpha < 0 < n$. Therefore, by the Measure Theoretic Finite-Dimensional Parametric Transversality Theorem (Appendix A), $\rho''(y)$ is transversal to $L_I \times \Delta_I$ for a set of full Lebesgue measure of y in $\tilde{N}$.

The dimension of $\tilde{T}M$ is $2m-1$ and that of $L_I \times \Delta_I$ is $\alpha + dm - (d-\alpha)m = (m+1)\alpha$. By (6.2.1), $2m - 1 < (d-\alpha)m + \alpha$ and hence $\dim \tilde{T}M + \dim L_I \times \Delta_I = 2m - 1 + (m+1)\alpha < (d-\alpha)m + \alpha + (m+1)\alpha = 2\alpha + dm = \dim T\mathbb{R}^\gamma \times M^d$. Hence, as usual, if the image of $\tilde{T}M$ intersected $L_I \times \Delta_I$, the intersection could not be transversal. Thus for a residual set of y, the image of $\tilde{T}M$ does not intersect $L_I \times \Delta_I$. Hence, either $T\Phi_{f,\varphi,I,y}(v) \notin L_I$ or $\tau(f)(x, y) \notin \Delta_I$. But for every $(x, y) \in M \times \tilde{N}$, we have $\tau(f)(x, y) \in \Delta_I$ for some I, and hence $T\Phi_{f,\varphi,I,y}(v) \notin L_I$ for a residual set of y, for all $v \in T_x M$, for all $x \in M$ for some I. But if $T_x\Phi_{f,\varphi,I,y}(v) \notin L_I$, then $T_x\Phi_{f,\varphi,y}(v) \neq 0$ and hence $\Phi_{f,\varphi,y}$ is immersive, as required.

6.3. *Injectivity of* Φ

As in Section 5, we aim to make $\Phi_{f,\varphi} \times \Phi_{f,\varphi}$ transversal to the diagonal $\hat{\Delta}$ in $\mathbb{R}^d \times \mathbb{R}^d$. Let $R = \{(i_1, i_1'), (i_2, i_2'), \dots, (i_{\beta_R}, i_{\beta_R}')\}$ be a subset of $J_I \times J_I$ (possibly empty) and as before define

$$J_{I,R} = \{i \in J_I\colon (i,i') \notin R \text{ for any } i' \in J_I\}.$$

Note that, unlike in Section 5, for the moment we cannot exclude the case $J_{I,R} = \varnothing$. Let γ' be the number of elements in $J_{I,R}$ and note that $\gamma' \geq \alpha_R = \alpha - \beta_R$. If $\gamma' > 0$, write $J_{I,R} = \{j'_1, j'_2, \dots, j'_{\gamma'}\}$, with $j'_1 < j'_2 < \cdots < j'_{\gamma}$, and as in Section 5.1, define the maps $\Phi_{f,\varphi,I,R}\colon M \times N \to \mathbb{R}^{\gamma'}$ and $\Phi_{\varphi,y,I,R}\colon M \times \{y\} \to \mathbb{R}^{\gamma'}$ by

$$\begin{aligned}\Phi_{f,\varphi,I,R}(x,y) &= (\varphi(x_{j'_1}), \varphi(x_{j'_2}), \dots, \varphi(x_{j'_{\gamma'}}))^{\dagger},\\ \Phi_{f,\varphi,I,R,y}(x) &= \Phi_{f,\varphi,I,R,y}(x,y).\end{aligned}$$

If $\gamma' = 0$ define $\Phi_{\varphi,y,I,R}\colon M \times \{y\} \to \mathbb{R}^0 = \{0\}$ by $\Phi_{f,\varphi,I,R}(x,y) = 0$ for all x, y. Observe that if $\Phi_{f,\varphi,I,R}(x,y) \neq \Phi_{f,\varphi,I,R}(x',y)$ then $\Phi_{f,g,\varphi}(x,y) \neq \Phi_{f,g,\varphi}(x',y)$. Thus if for all $(x,x') \in (M \times M)\backslash\Delta$, we have $\Phi_{f,\varphi,I,R,y}(x) \neq \Phi_{f,\varphi,I,R,y}(x')$ for some I, R, then $\Phi_{f,\varphi,y}$ is injective on M.

Define $\sigma\colon \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M,\mathbb{R}) \to \mathcal{C}^n((M \times M)\backslash\Delta \times \tilde{N}, \mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'})$ and $\tau\colon \mathcal{D}^r(M \times N, M) \to \mathcal{C}^n((M \times M)\backslash\Delta \times \tilde{N}, M^d \times M^d)$ by

$$\begin{aligned}\sigma(f,\varphi)(x,x',y) &= (\Phi_{\varphi,I,R}(x,y), \Phi_{\varphi,I,R}(x',y)),\\ \tau(f)(x,x',y) &= ((x, f^{(1)}(x,y)\dots, f^{(d-1)}(x,y)), (x', f^{(1)}(x',y)\dots, f^{(d-1)}(x',y))),\end{aligned}$$

and $\sigma'\colon \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M,\mathbb{R}) \to \mathcal{C}^n((M \times M)\backslash\Delta \times \tilde{N}, \mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'} \times M^d \times M^d)$ by

$$\sigma'(f,\varphi) = (\sigma(f,\varphi), \tau(f)).$$

By Corollaries C.5 and C.6, ev_τ and ev_σ are $\mathcal{C}^n$ (since $r \geq 2n$). From Section 5.6, recall that $\Delta_{I,R}$ is a codimension $(d - \alpha_R)m$ submanifold of $M^d \times M^d$ defined by

$$\begin{aligned}\Delta_{I,R} = \{&(z_0,\dots,z_{d-1},z'_0,\dots,z'_{d-1}) \in M^d \times M^d\colon (z_0,\dots,z_{d-1}) \in \Delta_I, \text{ and}\\ &\text{if } (i,i') \in J_I \times J_I \text{ then } z_i = z'_{i'} \text{ if and only if } (i,i') \in R\},\end{aligned}$$

and let $\hat{\Delta}_{\gamma'}$ be the diagonal in $\mathbb{R}^{\gamma''} \times \mathbb{R}^{\gamma''}$.

Proposition 6.2. *The map $ev_{\sigma'}$ is transversal to $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$ for all I, R.*

Proof. Suppose that $ev_{\sigma'}(\Xi) \in \hat{\Delta}_{\gamma''} \times \Delta_{I,R}$ for some $\Xi = (f,\varphi,x,x',y)$, with $(x,x',y) \in (M \times M\backslash\Delta) \times \tilde{N}$. First we show that if $\tau(f)(x,x',y) \in \Delta_{I,R}$, then given any $(\bar{u},\bar{u}') = ((u_0,u_1,\dots,u_{d-1}),(u'_0,u'_1,\dots,u'_{d-1})) \in ((T_{x_0}M \times \cdots \times T_{x_{d-1}}M) \times (T_{x'_0}M \times \cdots \times T_{x'_{d-1}}M))$ we can find a $\eta_0 \in T_f\mathcal{D}^r(M \times N, M)$ such that $T_{f,x,x',y}(ev_\tau)(\eta_0,u_0,u'_0,0) = (\bar{u},\bar{u}')$. By Corollary C.5 we have

$$T_{f,x,x',y}(ev_\tau) = ((\eta_0(x,y),\dots,\eta_{d-1}(x,y)), (\eta'_0(x',y),\dots,\eta'_{d-1}(x',y))),$$

where η_i and η'_i satisfy

$$\begin{aligned}\eta_{i+1}(x,y) &= \eta(x_i,y_i) + T_{(x_i,y_i)}f(\eta_i(x,y),0),\\ \eta'_{i+1}(x',y) &= \eta(x'_i,y_i) + T_{(x'_i,y_i)}f(\eta'_i(x',y),0),\end{aligned}$$

with of course $\eta_0 = \eta'_0 = 0$. For $y \in \tilde{N}$ the points $\{y_i\colon\ i = 0, \ldots, d-1\}$ are disjoint, and hence $(x_i, y_i) \neq (x_{i'}, y_{i'})$, $(x'_i, y_i) \neq (x'_{i'}, y_{i'})$ and $(x_i, y_i) \neq (x'_{i'}, y_{i'})$ for all $i \neq i'$. Furthermore since $x \neq x'$ (recall we only consider $(x, x') \in (M \times M \backslash \Delta)$) and (f, g) is a diffeomorphism, we have $(x_i, y_i) \neq (x'_i, y_i)$ for all i. Hence the points $\{(x_i, y_i)\colon\ i = 0, \ldots, d-1\} \cup \{(x'_i, y_i)\colon\ i = 0, \ldots, d-1\}$ are all disjoint. Thus, using Corollary C.12 as usual, we can find a $\eta \in T_f \mathcal{D}^r(M \times N, M)$ such that $\eta(x_i, y_i)$ for $i = 0, \ldots, d-1$ take on whatever values we want, and such that $\eta(x'_i, y_i) = 0$ for $i = 0, \ldots, d-1$. Then, as in Lemma 5.12, given any $\bar{u} \in (T_{x_0}M \times \cdots \times T_{x_{d-1}}M)$ we can find a η such that $T_{f,x,x',y}(ev_\tau)(\eta, u_0, 0_{x'}, 0_y) = (\bar{u}, 0_{\bar{x}'})$ where $0_{\bar{x}'} = (0_{x'_0}, \ldots, 0_{x'_{d-1}})$. Similarly, given $\bar{u}' \in (T_{x'_0}M \times \cdots \times T_{x'_{d-1}}M)$ we can find a η' such that $T_{f,x,x',y}(ev_\tau)(\eta', 0_x, u'_0, 0_y) = (0_{\bar{x}}, \bar{u}')$, where $0 = (0_{x_0}, \ldots, 0_{x_{d-1}})$. Hence $T_{f,x,x',y}(ev_\tau)(\eta + \eta', u_0, u'_0, 0_y) = (\bar{u}, \bar{u}')$, as required.

Now let us turn to the other component of σ'. If $\gamma' = 0$, this is trivial, and hence the surjectivity of $T_{f,x,x',y}(ev_\tau)$ is sufficient to ensure the transversality of $ev_{\sigma'}$. If $\gamma' > 0$, denote $ev_{\sigma'}(\Xi) = (z, z, \bar{x}, \bar{x}') \in \hat{\Delta}_{\gamma'} \times \Delta_{I,R}$, where $\bar{x} = (x_0, x_1, \ldots, x_{d-1})$ and $\bar{x}' = (x'_0, x'_1, \ldots, x'_{d-1})$. By linearity we have

$$
\begin{aligned}
T_\Xi(ev'_\sigma)(\eta, \xi, u_0, u'_0, 0_y) &= T_\Xi(ev_{\sigma'})(\eta, 0_\varphi, u_0, u'_0, 0_y) + T_\Xi(ev_{\sigma'})(0_f, \xi, 0_x, 0_{x'}, 0_y) \\
&= (\bar{w}_0, \bar{w}'_0, \bar{u}, \bar{u}') + T_\Xi(ev_{\sigma'})(0_f, \xi, 0_x, 0_{x'}, 0_y),
\end{aligned}
$$

for some $(\bar{w}_0, \bar{w}'_0) \in T_{(z,z)}(\mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'})$. By definition

$$
\begin{aligned}
T_\Xi(ev_{\sigma'})(0_f, \xi, 0_x, 0_{x'}, 0_y) &= (T_\Xi(ev_\sigma)(0_f, \xi, 0_x, 0_{x'}, 0_y), T_{f,x,x',y}(ev_\tau)(0_f, 0_x, 0_{x'}, 0_y)) \\
&= (T_\Xi(ev_\sigma)(0_f, \xi, 0_x, 0_{x'}, 0_y), (0_{\bar{x}}, 0_{\bar{x}})).
\end{aligned}
$$

Finally, as in Section 5.8, we have

$$
T_\Xi(ev_\sigma)(0_f, \xi, 0_x, 0_{x'}, 0_y) = \left(\begin{pmatrix} \xi(x_{j'_1}) \\ \xi(x_{j'_2}) \\ \vdots \\ \xi(x_{j'_{\gamma''}}) \end{pmatrix}, \begin{pmatrix} \xi(x'_{j'_1}) \\ \xi(x'_{j'_2}) \\ \vdots \\ \xi(x'_{j'_{\gamma''}}) \end{pmatrix} \right).
$$

Since $\tau(f)(x, x', y) \in \Delta_{I,R}$ the points $x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}$ are distinct and $\{x_{j'_1}, x_{j'_2}, \ldots, x_{j'_{\gamma'}}\} \cap \{x'_{j'_1}, x'_{j'_2}, \ldots, x'_{j'_{\gamma'}}\} = \varnothing$. Thus, using Corollary C.12, given any $(\bar{w}, \bar{w}') \in T_z\mathbb{R}^{\gamma'} \times T_z\mathbb{R}^{\gamma'}$ we can find a $\xi \in T_\varphi(\mathcal{C}^r(M, \mathbb{R}))$ such that $T_\Xi(ev_\sigma)(0_f, \xi, 0_x, 0_{x'}, 0_y) = ((\bar{w} - \bar{w}_0 - (\bar{w}' - \bar{w}'_0)), 0_z)$. Thus

$$
\begin{aligned}
T_\Xi(ev_{\sigma'})(\eta, \xi, u_0, u_{0'}, 0_y) &= (\bar{w}_0, \bar{w}'_0, \bar{u}, \bar{u}') + ((\bar{w} - \bar{w}_0 - (\bar{w}' - \bar{w}'_0)), 0_z, (0_{\bar{x}}, 0_{\bar{x}'})) \\
&= ((\bar{w} - \bar{w}' + \bar{w}'_0), \bar{w}'_0, \bar{u}, \bar{u}').
\end{aligned}
$$

But $(\bar{w}' - \bar{w}'_0, \bar{w}' - \bar{w}'_0) \in T_{(z,z)}\hat{\Delta}_{\gamma''}$, and hence $(\bar{w}, \bar{w}', \bar{u}, \bar{u}') \in \text{Image}(T_\Xi(ev_{\sigma'})) + T_{(z,z)}\hat{\Delta}_{\gamma'}$, as required. □

We have $\dim\,(M \times M)\backslash\Delta \times \tilde{N} - \operatorname{codim} \hat{\Delta}_{\gamma'} \times \Delta_{I,R} = 2m + n - \gamma' - (d - \alpha_R)m$. Since $\gamma' \geq \alpha_R$, we get

$$\begin{aligned} 2m + n - \gamma' - (d-\alpha_R)m &\leq d - 1 - (d - \alpha_R)m - \gamma' + n \\ &\leq (d - \alpha_R)(1 - m) - 1 + n \\ &< n. \end{aligned} \tag{6.3.1}$$

Since $ev_{\sigma'}$ is $\mathcal{C}^n$, the Parametric Transversality Theorem implies that there is a residual set of $(f, \varphi) \in \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$ for which $\sigma'(f, \varphi)$ is transversal to $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$. Fix any such (f, φ), and define $\sigma''\colon\ \tilde{N} \to \mathcal{C}^n((M \times M)\backslash\Delta, \mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'} \times M^d \times M^d)$ by

$$\begin{aligned} \sigma''(y)(x, x') &= \sigma'(f, \varphi)(x, x', y) \\ &= (\Phi_{f,\varphi,I,R,y}(x), \Phi_{f,\varphi,I,R,y}(x'), \tau(f)(x, x', y)). \end{aligned}$$

Then $ev_{\sigma''}(y, x, x') = \sigma'(f, \varphi)(x, x', y)$, and hence $ev_{\sigma''}$ is transversal to $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$. Also $ev_{\sigma''}$ is $\mathcal{C}^n$, and using (6.3.1), we have $\dim\,(M \times M)\backslash\Delta - \operatorname{codim} \hat{\Delta}_{\gamma'} \times \Delta_{I,R} < 0$. Thus by the Measure Theoretic Finite-Dimensional Parametric Transversality Theorem, there is a set of y of full Lebesgue measure in $\tilde{N}$ such that $\sigma''(y)$ is transversal to $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$. The dimension of $(M \times M)\backslash\Delta$ is $2m$ and that of $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$ is $\gamma' + 2dm - (d - \alpha_R)m$. By (6.3.1) we have $2m - (d - \alpha_R)m < \gamma'$ and hence $\dim(M \times M)\backslash\Delta + \dim \hat{\Delta}_{\gamma'} \times \Delta_{I,R} = 2m + \gamma' + 2dm - (d - \alpha_R)m < 2\gamma' + 2dm = \dim \mathbb{R}^{\gamma'} \times \mathbb{R}^{\gamma'} \times M^d \times M^d$. Thus, as usual, if the image of $(M \times M)\backslash\Delta$ intersected $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$, the intersection could not be transversal. Thus for a residual set of y, the image of $(M \times M)\backslash\Delta$ does not intersect $\hat{\Delta}_{\gamma'} \times \Delta_{I,R}$. Hence, either $(\Phi_{f,\varphi,I,R,y}(x), \Phi_{f,\varphi,I,R,y}(x')) \notin \hat{\Delta}_{\gamma'}$ or $\tau(f)(x, x', y) \notin \Delta_{I,R}$. Note, incidentally, that if $\gamma' = 0$ then by definition $(\Phi_{f,\varphi,I,R,y}(x), \Phi_{f,\varphi,I,R,y}(x')) = (0, 0) \in \hat{\Delta}_{\gamma'}$, and hence for our choice of f we have $\tau(f)(x, x', y) \notin \Delta_{I,R}$ for all I, R such that $\gamma' = 0$. Hence although we could not a priori exclude this case, it does not in fact arise.

To conclude the proof, we note that for every $(x, x', y) \in ((M \times M)\backslash\Delta) \times \tilde{N}$ we have $\tau(f)(x, x', y) \in \Delta_{I,R}$ for some I and R, and hence $(\Phi_{f,\varphi,I,R,y}(x), \Phi_{f,\varphi,I,R,y}(x')) \notin \hat{\Delta}_{\gamma'}$ for some I and R. Thus for a residual set of y, $\Phi_{f,g,\varphi,y}(x) \neq \Phi_{f,g,\varphi,y}(x')$ for all $(x, x') \in (M \times M)\backslash\Delta$, as required.

6.4. Lower Degrees of Smoothness

Finally, we show how to deduce Theorem 3.2 for $1 \leq r < 2n$, from the above proof for $r \geq 2n$. Fix r such that $1 \leq r < 2n$ and denote $\mathcal{B}^r = \mathcal{D}^r(M \times N, M) \times \mathcal{C}^r(M, \mathbb{R})$. For $(f, \varphi) \in \mathcal{B}^r$ let $\tilde{N}(f, \varphi) \subset N$ be the set of y such that $\Phi_{f,g,\varphi,y}$ is an embedding, and let $B(f, \varphi) = N\backslash\tilde{N}(f, \varphi)$. Note that since $\Phi_{f,g,\varphi,y}$ depends continuously on y and embeddings are open in $\mathcal{C}^r(M, \mathbb{R}^d)$, the set $\tilde{N}(f, \varphi)$ is necessarily open. Given an $\varepsilon > 0$, define the ε-neighbourhood of $B(f, \varphi)$ by $B(f, \varphi, \varepsilon) = \{y \in N\colon\ d(y, B(f, \varphi)) < \varepsilon\}$.

Let $\mathcal{E}^r$ be the set of (f, φ) in $\mathcal{B}^r$ for which $\mu(\tilde{N}(f, \varphi)) = 1$, where μ is Lebesgue measure. By the above $\mathcal{E}^{2n}$ is dense in $\mathcal{B}^{2n}$. Since the latter is dense in $\mathcal{B}^r$ we see that $\mathcal{E}^{2n}$ is dense in $\mathcal{B}^r$. But $\mathcal{E}^{2n} \subset \mathcal{E}^r$, and hence $\mathcal{E}^r$ is dense in $\mathcal{B}^r$. This space is separable and hence we may choose a countable set $\{(f_i, \varphi_i) \in \mathcal{E}^r\colon\ i \in \mathbb{N}\}$ that is dense in $\mathcal{B}^r$. For any i, we have

$$B(f_i, \varphi_i) = \bigcap_{\varepsilon>0} B(f_i, \varphi_i, \varepsilon),$$

and since $\mu(B(f_i, \varphi_i)) = 0$, we have $\mu(B(f_i, \varphi_i, \varepsilon)) \to 0$ as $\varepsilon \to 0$. Hence, given $\delta > 0$, we may choose an $\varepsilon(i, \delta) > 0$ such that $\mu(N \backslash B(f_i, \varphi_i, \varepsilon(i, \delta))) > 1 - \delta$. Now, $N \backslash B(f_i, \varphi_i, \varepsilon)$ is closed and hence compact. Using the continuity of $\Phi_{f,g,\varphi,y}$ with respect to f, φ, and y and the density of embeddings in $\mathcal{C}^r(M, \mathbb{R}^d)$, we can find an open neighbourhood $\mathcal{N}(f_i, \varphi_i, \delta)$ of (f_i, φ_i) in $\mathcal{B}^r$ such that $\Phi_{f,g,\varphi,y}$ is an embedding for all $(f, \varphi) \in \mathcal{N}(f_i, \varphi_i, \delta)$ and $y \in N \backslash B(f_i, \varphi_i, \varepsilon(i, \delta))$. Since $\{(f_i, \varphi_i): i \in \mathbb{N}\}$ is dense in $\mathcal{B}^r$, the union of $\mathcal{N}(f_i, \varphi_i, \delta)$ is open and dense in $\mathcal{B}^r$. Thus if we let δ_n be a sequence such that $\delta_n \to 0$ and define

$$\mathcal{N} = \bigcap_{n\in\mathbb{N}} \bigcup_{i\in\mathbb{N}} \mathcal{N}(f_i, \varphi_i, \delta_n),$$

we see that $\mathcal{N}$ is residual. Now if $(f, \varphi) \in \mathcal{N}$, then there exists a sequence i_n such that $(f, \varphi) \in \mathcal{N}(f_{i_n}, \varphi_{i_n}, \delta_n)$ for each n. Hence $\Phi_{f,g,\varphi,y}$ is an embedding for all $y \in N \backslash B(f_{i_n}, \varphi_{i_n}, \varepsilon(i, \delta_n))$ for each n. Hence if we define

$$\tilde{N}'(f, \varphi) = \bigcup_{n\in\mathbb{N}} N \backslash B(f_{i_n}, \varphi_{i_n}, \varepsilon(i, \delta_n)),$$

then $\tilde{N}'(f, \varphi) \subset \tilde{N}(f, \varphi)$. Since $\mu(N \backslash B(f_{i_n}, \varphi_{i_n}, \varepsilon(i, \delta_n))) \to 1$ as $n \to \infty$, we have $\mu(\tilde{N}'(f, \varphi)) = 1$ and hence $\mu(\tilde{N}(f, \varphi)) = 1$. Thus $(f, \varphi) \in \mathcal{E}^r$, and so $\mathcal{N} \subset \mathcal{E}^r$. Therefore $\mathcal{E}^r$ contains a residual set, as required.

Appendix A. Transversality

If M and N are finite-dimensional manifolds, $L \subset N$ a submanifold and $g: M \to N$ a smooth map, we say that g is *transversal* to L (written as $g \pitchfork L$) if for all $x \in g^{-1}(L)$ the tangent space $T_{g(x)}N$ at $g(x)$ is spanned by the image $T_x g(T_x M)$ of the tangent space at x under the derivative of g and the tangent space $T_{g(x)}L$ to L, i.e.,

$$T_{g(x)}N = T_{g(x)}L + T_x g(T_x M). \tag{A1}$$

This sum need not be direct and no condition is imposed on $T_x g(T_x M)$ if $g(x) \notin L$. If (A1) holds only for all $x \in g^{-1}(L) \cap K$ for some subset $K \subset M$ we say that g is transversal to L on, or along, K (written as $g \pitchfork_K L$).

One important consequence of transversality is that if $g \pitchfork L$, then $g^{-1}(L)$ is a submanifold of M ([Thom, 1954 and 1956]). This is a generalization of the well-known fact that if $T_x g$ is surjective (so that $g(x)$ is a regular value of g), then $g^{-1}(g(x))$ is a submanifold of M.

The notion of transversality gives a formal meaning to the notion of “typical intersection,” as in statements such as “a line and a surface in $\mathbb{R}^3$ typically meet in isolated points.” The theorem that lies behind such intuition is Thom’s Transversality Theorem.

Transversality Theorem. [Thom, 1954 and 1956] *The set of g which are transversal to $L \subset N$ is residual in $\mathcal{C}^r(M, N)$, $r \geq 1$.*

This is a straightforward consequence of the Morse-Sard Theorem, which says that for any $g \in \mathcal{C}^r(M, N)$, with $r \geq \max\{0, \dim M - \dim N\}$, the set of regular values of g (points $g(x)$ for x such that $T_x g$ is surjective) is residual in N.

The notion of transversality can be extended to infinite-dimensional manifolds (modelled on Banach spaces). If either M or N is finite-dimensional, then we can use condition (A1) unchanged, but if both are infinite-dimensional then we also require "splitting' conditions on the kernel and image of $T_x g$. More precisely, we say that g is transversal to $L \subset N$ if for all $x \in g^{-1}(L)$ we have

$$(T_x g)^{-1}(T_{g(x)} L) \text{ has a closed complement in } T_x M, \tag{A2a}$$

$$T_x g(T_x M) \text{ has a closed subspace } V \text{ such that } T_{g(x)} N = T_{g(x)} L + V. \tag{A2b}$$

The reason that these additional conditions are not required when at least one of M or N is finite-dimensional is that finite dimensional subspaces are always closed and have closed complements, and $(T_x g)^{-1}$ of a closed subspace is closed. This will always be the setting in which we employ transversality in this paper, and hence conditions A2 never need to be verified explicitly.

One of the most useful applications of the notion of transversality in infinite dimensions is Abraham's Parametric Transversality Theorem (Abraham [1963], Abraham and Robbin [1967], Quinn [1970]; also see Hirsch [1976] and Palis and de Melo [1982] for the case where A is finite-dimensional). We make extensive use of this theorem in this paper, in the form given in Abraham and Robbin [1967]:

Parametric Transversality Theorem. *Let $\mathcal{A}$, M, and N be $\mathcal{C}^r$ manifolds and ρ: $\mathcal{A} \to \mathcal{C}^r(M, N)$ be a map such that the evaluation map ev_ρ: $\mathcal{A} \times M \to N$ given by $ev_\rho(a, x) = \rho(a)(x)$ is $\mathcal{C}^r$. Let $L \subset N$ be a submanifold of finite codimension p in N. Suppose that $\mathcal{A}$ and M are second countable (i.e., their topology has a countable base), that M has finite dimension m with $r > \max\{0, m - p\}$, and that ev_ρ is transversal to L. Then the set of a such that $\rho(a)$ is transversal to L is residual in $\mathcal{A}$. Furthermore if L is closed and M is compact, then the set of such a is open.*

The significance of this may perhaps best be illustrated if we let $L = \{y \in N\}$ be a single point. The theorem then says that if perturbations in both a and x (where $y = \rho(a)(x)$) are able to generate the whole of $T_y N$, then for a generic a perturbations in x alone give $T_y N$, i.e., $T_x \rho(a)$ is surjective. For a more general submanifold L, we have the statement that if perturbations in both a and x are able to generate a complement to $T_y L$ in $T_y N$, then for a generic a it is sufficient to consider perturbations in x alone to obtain this complement.

The proof of the Parametric Transversality Theorem is based upon Smale's Density Theorem (Smale [1965]), which is an infinite-dimensional generalization of the Morse-Sard Theorem. The underlying idea is very simple: Since ev_ρ is transversal to L, the set $(ev_\rho)^{-1}(L)$ is a submanifold of $\mathcal{A} \times M$. Let π: $\mathcal{A} \times M \to \mathcal{A}$ be the projection onto the first factor. A simple unravelling of the definitions shows that $\rho(a)$ is transversal to L if and only if a is a regular value of π restricted to $(ev_\rho)^{-1}(L)$. But by Smale's Density Theorem the set of such regular values is residual in $\mathcal{A}$. Note that if $\mathcal{A}$ is in fact finite-dimensional we can instead apply the standard Morse-Sard Theorem to obtain the following:

Measure Theoretic Finite-Dimensional Parametric Transversality Theorem. *Let $\mathcal{A}$, M, and N be $\mathcal{C}^r$ manifolds and ρ: $\mathcal{A} \to \mathcal{C}^r(M, N)$ be a map such that ev_ρ is $\mathcal{C}^r$. Let $L \subset N$ be a submanifold of finite codimension p in N. Suppose that $\mathcal{A}$ and M are finite-dimensional, that $r > \max\{0, m - p\}$ where m is the dimension of M and that ev_ρ is transversal to L. Then the set of a such that $\rho(a)$ is not transversal to L has zero Lebesgue measure in $\mathcal{A}$. Furthermore if L is closed and M is compact, then the set of such a is open.*

As far as we are aware, there is no published statement of this version.

Recall (e.g., Abraham [1963], Eells [1966], Eliasson [1967], Foster [1975], Franks [1979], and see Appendix C below) that if M is compact and N is finite-dimensional, then $\mathcal{C}^r(M, N)$ for any $0 \leq r < \infty$ has the structure of a C^∞ Banach manifold. Taking $\mathcal{A} = \mathcal{C}^r(M, N)$, with ρ the identity (giving the trivial evaluation map $ev_\rho(f, x) = f(x)$), we see that the Parametric Transversality Theorem immediately implies Thom's original Transversality Theorem.

A further important property of transversal intersection is that in some sense it is structurally stable. This is best expressed by the following.

Transversal Isotopy Theorem. (Abraham and Robbin, [1967]) *Let $\mathcal{A}$, M, N, L, and ρ be as in the Parametric Transversality Theorem. Suppose that L is closed and M is $\mathcal{C}^{r+2}$ and compact. Suppose that for some $a_0 \in \mathcal{A}$, $\rho(a_0)$ is transversal to L. Then there is an open neighbourhood U of a_0 in $\mathcal{A}$ such that for all $a \in U$, $\rho(a)$ is transversal to L and $\rho(a)^{-1}(L)$ is $\mathcal{C}^{r-1}$ isotopic to $\rho(a_0)^{-1}(L)$. This means that there is a $\mathcal{C}^{r-1}$ map h: $M \times [0, 1] \to M$ such that $h(\bullet, t)$ is a diffeomorphism of M for all $t \in [0, 1]$, $h(\bullet, 0) = Id$, and $h(\bullet, 1)$ maps $\rho(a)^{-1}(L)$ onto $\rho(a_0)^{-1}(L)$. In particular $\rho(a)^{-1}(L)$ and $\rho(a_0)^{-1}(L)$ are diffeomorphic.*

Appendix B. Geometry of Function Spaces

In applying the Parametric Transversality Theorem, the first step is always to show that an appropriate function space is in fact a smooth manifold, and that the appropriate evaluation function is sufficiently smooth. In this appendix we collect together a number of the underlying results that permit us to do this.

Much of this material requires familiarity with more advanced concepts of differential geometry, such as connections and exponential maps, than the remainder of the paper. It is rather unfortunate that these are used extensively in the construction of the manifold structure on function spaces and in the analysis of the composition and tangent operators, even though the subsequent results can be stated without reference to such terms. Thus for instance whilst a Riemannian structure appears necessary to construct a manifold structure on $\mathcal{C}^r(M, N)$, once constructed this is independent of the choice of original Riemannian structure. Once we have proved the results in this appendix therefore, we need not make use of such differential geometric tools elsewhere in the paper, and a reader unfamiliar with such concepts can safely skip most of this appendix.

B.1. Function Spaces as Manifolds

It has long been known that if M is compact, then subject to some mild conditions on N, the space $\mathcal{C}^r(M, N)$ for any $0 \leq r < \infty$ has the structure of a Banach manifold. The earliest such results are due to Eells [1958] and Palais [1968]. Since then, many extensions and generalizations have appeared (e.g., see Eells [1966], Eliasson [1967], Foster [1975], Franks [1979]). For the purposes of this paper, the version due to Eliasson is the most useful.

Theorem B.1. (Eliasson [1967]) *Suppose that M is a compact $\mathcal{C}^\infty$ Riemannian manifold and N is a paracompact $\mathcal{C}^\infty$ Banach manifold, without boundary, admitting a $\mathcal{C}^\infty$ connection and a Finsler structure. Then $\mathcal{C}^r(M, N)$ for any $0 \leq r < \infty$ is a paracompact $\mathcal{C}^\infty$ Banach manifold that admits a $\mathcal{C}^\infty$ connection and a Finsler structure. The manifold structure on $\mathcal{C}^r(M, N)$ is independent of the choice of Riemannian structure on M and N.*

In order to describe this structure on $\mathcal{C}^r(M, N)$, recall (e.g., [Hirsch, 1976]) that given a map $f \in \mathcal{C}^r(M, N)$, the pull-back bundle $f^*_{\tau_N}\colon f^*TN \to M$ is the vector bundle given by $f^*TN = \{(x, v) \in M \times TM\colon f(x) = \tau_N(v)\}$ and $f^*_{\tau_N}(x, v) = x$, where $\tau_N\colon TN \to N$ is the tangent bundle of N. Informally, f^*TN is the bundle over M whose fibre $(f^*TN)_x$ over $x \in M$ is $T_{f(x)}N$. The space of $\mathcal{C}^r$ sections of f^*TN, endowed with the $\mathcal{C}^r$ norm, is a Banach space which we denote $\mathcal{C}^r(f^*TN)$. This turns out to be the tangent space $T_f\mathcal{C}^r(M, N)$ of $\mathcal{C}^r(M, N)$ at f. The union over all f of the $T_f\mathcal{C}^r(M, N)$ is thus $\mathcal{C}^r(M, TN)$. This is given the structure of the tangent bundle of $\mathcal{C}^r(M, N)$ by $\tau_\mathcal{C}\colon \mathcal{C}^r(M, TN) \to \mathcal{C}^r(M, N)$ defined by $\tau_\mathcal{C}(\eta) = \tau_N \circ \eta$.

To describe how $\mathcal{C}^r(M, N)$ is a manifold modelled on $\mathcal{C}^r(f^*TN)$, first note that if $\eta \in \mathcal{C}^r(f^*TN)$, then $\eta(x) = (x, \tilde{\eta}(x))$ for some map $\tilde{\eta}\colon M \to TN$ such that $\tilde{\eta}(x) \in T_{f(x)}N$ for all $x \in M$. It shall be convenient to identify η and $\tilde{\eta}$ and hence think of the space of sections of f^*TN as $\mathcal{C}^r(f^*TN) = \{\eta \in \mathcal{C}^r(M, TN)\colon \tau_N \circ \eta = f\}$. This allows us to define a chart $h_f\colon \mathcal{U}_f \to \mathcal{V}_f$ for $\mathcal{C}^r(M, N)$ centred at f, where $\mathcal{V}_f$ is a sufficiently small neighbourhood of the zero section in $\mathcal{C}^r(f^*TN)$, $\mathcal{U}_f = h_f^{-1}(\mathcal{V}_f)$, and h_f is defined by $h_f^{-1}(\eta) = \exp_N \circ \eta$, with $\exp_N\colon TN \to N$ the exponential map obtained from the connection on N (e.g., [Lang, 1972] or [Irwin, 1980]). Recall that if we denote the restriction of $\exp_N$ to the fibre over $f(x)$ by $\exp_{f(x)}\colon T_{f(x)}N \to N$, then $\exp_{f(x)}$ maps a sufficiently small neighbourhood of $0_{f(x)}$ diffeomorphically onto its image. This ensures that h_f^{-1} is a homeomorphism, with h_f itself given by $h_f(g)(x) = (\exp_{f(x)})^{-1}(g(x))$ for $g \in \mathcal{U}_f$. Observe that, as required, $\exp_N \circ \eta$ is indeed a $\mathcal{C}^r$ map from M to N. Furthermore, if 0_f is the zero section in $\mathcal{C}^r(f^*TN)$, so that $0_{f(x)} = 0_{f(x)}$, then $h_f^{-1}(0_f)(x) = \exp_N(0_{f(x)}) = \exp_N(0_{f(x)}) = f(x)$. Thus $h_f(f) = 0_f$.

To confirm that this indeed confers a manifold structure on $\mathcal{C}^r(M, N)$, it is necessary to verify that if $f, f' \in \mathcal{C}^r(M, N)$ with $\mathcal{U}_f \cap \mathcal{U}'_f \neq \varnothing$, then the transition map $h_{f'} \circ h_f^{-1}$ is $\mathcal{C}^\infty$

on $h_f(\mathcal{U}_f \cap \mathcal{U}_{f'})$. This follows (e.g., see [Foster, 1975]) from the smoothness of $\exp_N$ since $h_{f'} \circ h_f^{-1}(\eta) = (\exp_{f'})^{-1} \circ (Id, \exp_N \circ \eta)$, where $(\exp_{f'})^{-1}(x, y) = (\exp_{f'}(x))^{-1}(y)$.

Note that any $\mathcal{C}^k$ finite-dimensional manifold for $1 \leq k < \infty$ can be given a compatible $\mathcal{C}^\infty$ Riemannian structure (e.g., Hirsch [1976]) and hence admits a $\mathcal{C}^\infty$ connection and a Finsler structure. Any Hausdorff separable manifold is paracompact, and hence in particular if N is compact, or $N = \mathbb{R}^n$ for some n, then N is paracompact.

An immediate consequence of the above theorem is that if M is a smooth compact manifold, then $\mathcal{C}^r(M, M)$ and $\mathcal{C}^r(M, \mathbb{R})$ are Banach manifolds, for any $0 \leq r < \infty$. Also, it is well known (e.g., Eells [1966]) that as topological spaces they are separable. Since the set of diffeomorphisms $\mathcal{D}^r(M)$ is open in $\mathcal{C}^r(M, M)$, this itself is a separable manifold.

Finally, note that in the statement of the Parametric Transversality Theorem, it is not necessary for $\mathcal{C}^r(M, N)$ to be a manifold, and hence in particular, M need not be compact. In such a case, $\rho\colon \mathcal{A} \to \mathcal{C}^r(M, N)$ may simply be defined as a map, not even necessarily continuous. In particular, the theorem places no regularity conditions on ρ, but merely on ev_ρ. Thus, for instance, the theorem can be used in situations where M is an open submanifold of a compact manifold M', $\mathcal{A} = \mathcal{C}^r(M', N)$ and $\rho\colon \mathcal{C}^r(M', N) \to \mathcal{C}^r(M, N)$ involves the restriction of $f \in \mathcal{C}^r(M', N)$ to $f|_M \in \mathcal{C}^r(M, N)$. The simplest example is where ρ does nothing else, so that it is simply the natural inclusion $\mathcal{C}^r(M', N) \subset \mathcal{C}^r(M, N)$. The map $ev_\rho\colon \mathcal{C}^r(M', N) \times M \to N$ is then just $ev_\rho(f, x) = f(x)$. A more complex example, especially relevant to this paper, is for instance given by $ev_\rho(f, x) = \varphi(f^i(x))$ for some $\varphi \in \mathcal{C}^r(N, \mathbb{R})$.

B.2. Smoothness of the Composition Operator

The other essential component of our proofs of various versions of Takens Theorem will be the smoothness of the mapping $(f, \varphi) \mapsto \Phi_{f,\varphi}$. Observe that each component of this just consists of the composition of φ with some power of f. Its smoothness thus follows from the following well-known result, due in various forms to Eells [1966], Foster [1975], and Franks [1979].

Theorem B.2. *Suppose that M and N are compact $\mathcal{C}^\infty$ manifolds and L is a $\mathcal{C}^\infty$ manifold, modelled on a Hilbert space. Then the composition map $\sigma\colon \mathcal{C}^r(M, N) \times \mathcal{C}^{r+k}(N, L) \to \mathcal{C}^r(M, L)$ given by $\sigma(f, g) = g \circ f$ is $\mathcal{C}^k$, and*

$$T_{f,g}\sigma(\eta, \zeta) = \zeta \circ f + Tg \circ \eta.$$

It is important to note the loss of smoothness here; this is due to the term Tg. Thus in the case $k = 1$, if $\sigma(f, g)$ is to be in $\mathcal{C}^r(M, L)$, we must have $T_{f,g}\sigma(\eta, \zeta) \in \mathcal{C}^r((g \circ f)^* TL)$ and hence in particular $Tg \circ h$ must be $\mathcal{C}^r$, and hence g must be $\mathcal{C}^{r+1}$. Similarly for $k > 1$, the kth derivative of σ will require the kth derivative of g.

Taking $M = \{x\}$ to be a single point, and using the natural identification $N = \mathcal{C}^0(\{x\}, N)$ we get the obvious corollary:

Corollary B.3. *If* N *is compact then the evaluation map* $ev\colon N \times \mathcal{C}^k(N, L) \to L$ *given by* $ev(x, g) = g(x)$ *is* $\mathcal{C}^k$*, and*

$$T_{x,g}ev(v, \zeta) = \zeta(x) + T_x g(v).$$

B.3. The Tangent Map

To prove the immersivity of $\Phi_{f,\varphi}$ in various situations, we shall need to consider the map $(f, \varphi) \mapsto T\Phi_{f,\varphi}$, and in particular we shall need to compute the tangent map to the corresponding evaluation map. A crucial element in this calculation is the tangent map $T_f\sigma'$ to the map $\sigma'(f) = Tf$. To obtain this we are in effect differentiating the derivative operator. Note that this is very different from the second derivative of f, i.e., $T_f\sigma'$ and $T(Tf)$ are very different objects.

B.3.1. The Range of the Tangent Operator. Before we can begin to compute $T_f\sigma'$, we need to determine an appropriate domain and range for the operator σ'. Whilst the choice of $\mathcal{C}^r(M, N)$ as the domain of σ' is obvious, the selection of its range is more problematical. Thus, since $\sigma'(f)$ is a $\mathcal{C}^{r-1}$ map between TM and TN, we might be tempted to simply take $\mathcal{C}^{r-1}(TM, TN)$. Unfortunately, since TM is not compact, this space has no natural manifold structure and we have no way of discussing the differentiability of σ'.

Since Tf is linear on fibres, it is determined by its action on the unit tangent bundle $\tilde{T}M = \{v \in TM\colon \|v\| = 1\}$. One possible solution is therefore to use $\mathcal{C}^{r-1}(\tilde{T}M, TN)$ as the range of σ'. Since $\tilde{T}M$ is compact, this has a natural manifold structure. This is consistent with our use of σ' throughout the remainder of the paper, since in determining the immersivity of $\Phi_{f,\varphi}$ we only apply $T\Phi_{f,\varphi}$ to unit tangent vectors and hence regard the domain of $ev_{\sigma'}$ as $\tilde{T}M \times \mathcal{C}^r(M, N)$. This approach, however, has a serious drawback, namely the difficulty of composing maps in $\mathcal{C}^{r-1}(\tilde{T}M, TN)$ with those in $\mathcal{C}^{r-1}(\tilde{T}N, TL)$. In particular, if $F(v) = 0$ for any $v \in \tilde{T}M$, then there is no way of forming $G \circ F$.

A more natural and elegant approach is to make full use of the linearity of Tf and let the range of σ' be the space of vector bundle maps between TM and TN, that is, the space of maps $F\colon M \to N$ such that F is fibre preserving (i.e., $F(T_xM) \subset T_{f(x)}N$ for some $f \in \mathcal{C}^r(M, N)$) and linear on each fibre T_xM. For a fixed f we can think of the space of such maps as the Banach space of sections of the linear map bundle $\mathcal{L}(TM,\ f^*TN)$. Recall that this is defined as follows: If E and E' are vector bundles over M, then $\mathcal{L}(E, E')$ is the vector bundle whose fibre $\mathcal{L}(E, E')_x$ over $x \in M$ is just the space $\mathcal{L}(E_x, E'_x)$ of linear mappings from E_x to E'_x. A section $\tilde{F} \in \mathcal{C}^{r-1}(\mathcal{L}(TM,\ f^*TN))$ through $\mathcal{L}(TM,\ f^*TN)$ can be written as $\tilde{F}(x)(v) = (x, Fx(v))$ for some linear map $F_x\colon T_xM \to T_{f(x)}N$. Then $F\colon TM \to TN$ defined by $F(v) = F_{\tau_M(v)}(v)$ is a vector bundle map over f. Conversely such an F gives rise to a section $\tilde{F} \in \mathcal{C}^{r-1}(\mathcal{L}(TM,\ f^*TN))$ by $\tilde{F}(x)(v) = (x, F(v))$. We shall henceforth identify F and $\tilde{F}$. Note that we take the smoothness of the section F to be different from the smoothness of the underlying map f, which is precisely the case with Tf. Such "mixed smoothness" bundle maps are studied in Foster [1975] and Irwin [1980].

Following Eliasson [1967], we shall denote the union of the $\mathcal{C}^{r-1}(\mathcal{L}(TM,\ f^*TN))$ over all $f \in \mathcal{C}^r(M, N)$ by $\mathcal{C}^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(M, N) * TN))$ and use this as the range of σ'. For convenience, we shall denote this space as $\mathcal{VB}^{r-1}(TM, TN)$ and also use the notation $\mathcal{VB}_f^{r-1}(TM, TN)$ for $\mathcal{C}^{r-1}(\mathcal{L}(TM,\ f^*TN))$. Since $\mathcal{VB}^{r-1}(TM, TN)$ consists of sections of bundles over M, i.e., of maps whose domain is M, it is reasonable to expect that it can be given a manifold structure. In fact, Eliasson [1967] shows that $\mathcal{VB}^{r-1}(TM, TN)$ is naturally a vector bundle over $\mathcal{C}^r(M, N)$, where the fibre over $f \in \mathcal{C}^r(M, N)$ is just $\mathcal{VB}_f^{r-1}(TM, TN)$. Furthermore, he shows that σ' is a $\mathcal{C}^\infty$ section of this bundle. Our principal aim in the remainder of this section is to derive an expression for $T\sigma'$. To do this it is more convenient to use the manifold structure on $\mathcal{VB}^{r-1}(TM, TN)$ induced by its vector bundle structure, rather than the vector bundle structure itself. It also turns out to be preferable to use charts centred on Tf, rather than 0_f as is done by Eliasson.

Since, as discussed above, an element of $\mathcal{VB}^{r-1}(TM, TN)$ may be regarded as a map from TM to TN, we expect by analogy with Section B1 that the manifold structure $\mathcal{VB}^{r-1}(TM, TN)$ will be modelled on a suitable space of maps from TM into the tangent bundle $T(TN)$ of TN. To proceed further, we therefore have to describe the structure of $T(TN)$ in more detail. The reader may find it helpful to consult, for instance, Abraham and Robbin [1967] or Irwin [1980] for particularly clear accounts of this material.

B.3.2. Properties of $T(TN)$. We shall denote the standard vector bundle structure of $T(TN)$ over TN by $\tau_{TN}\colon T(TN) \to TN$. Suppose that $(T_U N, \alpha)$ is a local trivialization for TN, where U is an open neighbourhood in N, $T_U N = (\tau_N)^{-1}(U) = TU$ is the restriction of TN to U and $\alpha\colon T_U N \to U \times V$ is a local vector bundle isomorphism onto $U \times V$, where V is the Banach space on which N is modelled. Then $T\alpha$ is an isomorphism $T\alpha\colon T(T_U N) \to T(U \times V)$. Identifying $T(U \times V)$ with $TU \times TV$, using α to identify TU with $U \times V$, and then using the natural identification of TV with $V \times V$, we see that locally $T(TN)$ has the form $U \times V \times V \times V$, so that we may take as local coordinates (y, v_1, v_2, v_3) with $y \in N$, $v_1, v_2, v_3 \in V$. Note that contrary to appearances, this does not give $T(TN)$ the structure of a vector bundle over N; in particular, coordinate changes contain a bilinear term in the last (i.e., v_3) coordinate (e.g., see Abraham and Robbin, [1967]).

Observe that τ_N is a smooth map and hence we can take its tangent map $T\tau_N\colon T(TN) \to TN$. It turns out that this also gives $T(TN)$ a vector bundle structure over TN, which is not the same structure as $\tau_{TN}\colon T(TN) \to TN$. Thus in local coordinates we have $\tau_{TN}(y, v_1, v_2, v_3) = (y, v_1)$ and $T\tau_N(y, v_1, v_2, v_3) = (y, v_2)$. Note that $\tau_N \circ \tau_{TN} = \tau_N \circ T\tau_N$. When we need to distinguish between the two structures, we shall write $(T(TN), \tau_{TN})$ and $(T(TN), T\tau_N)$, or simply just τ_{TN} and $T\tau_N$. Fibres in the two respective structures can be denoted as $(\tau_{TN})^{-1}(u)$ and $(T\tau_N)^{-1}(u)$. When we write $T(TN)$ on its own, we always imply the standard structure τ_{TN}, and in particular we shall usually denote $(\tau_{TN})^{-1}(u)$ by $T_u(TN)$.

It turns out that the two structures are isomorphic, with a natural vector bundle isomorphism given by the so-called *canonical involution* $\omega\colon T(TN) \to T(TN)$. This satisfies $\tau_{TN} = T\tau_N \circ \omega$ and $\omega \circ \omega = Id$, and hence also $\tau_{TN} \circ \omega = T\tau_N$. In local coordinates it is given by $\omega(y, v_1, v_2, v_3) = (y, v_2, v_1, v_3)$. We shall make considerable use of ω throughout the paper.

Next, given a $u \in T_yN$, we wish to define the *vertical subspace* in $T_u(TN)$ as the space of $w \in T_u(TN)$ satisfying $T\tau_N(w) = 0_y$. In local coordinates we have $\{w \in T_u(TN)\colon\ T\tau_N(w) = 0_y\} = \{(y, \tilde{u}, 0, v_3) \in U \times V \times V \times V\}$, where $u = (y, \tilde{u})$. This suggests the following alternative characterization, which we shall make use of later:

Lemma B.4. $\{w \in T_u(TN)\colon\ T\tau_N(w) = 0_y\} = T_u(T_yN)$.

Proof. Since $\tau_N(u) = y$ for all $u \in T_yN$, we have$T\tau_N(w) = 0_y$ for all $w \in T(T_yN)$, and in particular $T\tau_N(T_u(T_yN)) = \{0_y\}$. Hence $T_u(T_yN) \subset \{w \in T_u(TN)\colon\ T\tau_N(w) = 0_y\}$. Conversely, if $w \in T_y(TN)$ but w is not tangent to T_yN, i.e., $w \notin T_u(T_yN)$, choose a smooth path $\gamma\colon\ [0, 1] \to TN$ such that $\gamma(0) = y$ and $T_0\gamma = w$. Then this path is not tangent to T_yN, and hence projecting down to N, we see that $\tau_N \circ \gamma$ is a smooth path in the neighbourhood of 0 which is not constant there, i.e. $T_0(\tau_N \circ \gamma) \neq 0_y$. Thus $T\tau_N(w) = T\tau_N(T_0\gamma) = T_0(\tau_N \circ \gamma) \neq 0_y$. Hence $\{w \in T_u(TN)\colon\ T\tau_N(w) = 0_y\} \subset T_u(T_yN)$, as required. □

Although $T_u(T_yN)$ is all that we require to describe the vector bundle structure on $\mathcal{VB}^{r-1}(TM, TN)$, in order to compute an expression for $T\sigma'$ it will also be necessary to define the spaces

$$W_{u,u'} = \{w \in T_u(TN)\colon\ T\tau_N(w) = u'\}.$$

Thus $T_u(T_yN) = W_{u,0}$, and $\omega(W_{u,u'}) = W_{u',u}$. Note that since $\tau_N \circ T\tau_N = \tau_N \circ \tau_{TN}$, we must have $\tau_N(u) = \tau_N(u')$ for this definition to make sense. In local coordinates $W_{u,u'} = \{(y, \tilde{u}, \tilde{u}', v_3) \in U \times V \times V \times V\}$, where $u = (y, \tilde{u})$ and $u' = (y, \tilde{u}')$. We thus see that $W_{u,u'}$ is nonempty for any $u, u' \in TN$ satisfying $\tau_N(u) = \tau_N(u')$. Also observe that whereas $T_u(T_yN)$ is a vector subspace of $T_u(TN)$, the space $W_{u,u'}$ for $u' \neq 0$ is not, since

Lemma B.5. *If* $w_1 \in W_{u,u_1}$, $w_2 \in W_{u,u_2}$, *and* $\lambda_1, \lambda_2 \in \mathbb{R}$, *then* $\lambda_1 w_1 + \lambda_2 w_2 \in W_{u,(\lambda_1 u_1 + \lambda_2 u_2)}$.

Proof. By definition, $T\tau_N$ is linear on fibres of τ_{TN}. Hence if $w_1, w_2 \in T_u(TN)$, we have $T\tau_N(\lambda_1 w_1 + \lambda_2 w_2) = \lambda_1 T\tau_N(w_1) + \lambda_2 T\tau_N(w_2) = \lambda_1 u_1 + \lambda_2 u_2$, as required. □

Finally, we turn to the exponential map on $T(TN)$. Recall that $\exp_N$ is a smooth map $\exp_N\colon\ TN \to N$, and hence $T\exp_N$ maps $T(TN)$ to TN. Note that by definition $T\exp_N$ is linear on fibres of τ_{TN}, i.e., on $T_u(TN)$ for any $u \in TN$. Since $\exp_y$ is a diffeomorphism on a neighbourhood of 0_y in T_yN, we immediately have

Lemma B.6. *For* $u \in T_yN$ *sufficiently small, the map* $T\exp_N$ *is a linear isomorphism between* $T_u(T_yN)$ *and* $T_{\exp(u)}N$.

Corollary B.7. *For* $u \in T_yN$ *sufficiently small, and any* $u' \in T_yN$, *given any* $v \in T_{\exp(u)}N$, *there exists a unique* $w \in W_{u,u'}$ *such that* $T\exp_N(\omega) = v$.

Proof. To show existence pick an arbitrary $w' \in W_{u,u'}$ (recall that this set is non-empty). Using Lemma B.6, choose $w_0 \in T_u(T_y N) = W_{u,0}$ such that $T\exp_N(w_0) = v - T\exp_N(w')$. Then, if we set $w = w_0 + w'$, we have $w \in W_{u,u'}$ by Lemma B.5. Furthermore, since $T\exp_N$ is linear on $T_u(TN)$, we have $T\exp_N(w) = T\exp_N(w_0) + T\exp_N(w') = v - T\exp_N(w') + T\exp_N(w') = v$, as required. Uniqueness follows similarly, for suppose that $w, w' \in W_{u,u'}$ with $T\exp_N(w) = T\exp_N(w') = v$. Then by Lemma B.5, $w - w' \in W_{u,0}$ and by the linearity of $T\exp_N$, we have $T\exp_N(w - w') = T\exp_N(w) - T\exp_N(w') = v - v = 0_{\exp(u)}$. Hence by Lemma B.6, $w - w' = 0$, as required. □

Denote the map from v to w by $\Psi_{u,u'}\colon T_{\exp(u)}N \to W_{u,u'}$. We have $T\exp_N \circ \Psi_{u,u'} = Id$, and since $T\exp_N$ is linear this gives the following "linearity" property:

Lemma B.8. *If $v_1, v_2 \in T_{\exp(u)}N$ and $\lambda_1, \lambda_2 \in \mathbb{R}$, then $\Psi_{u,(\lambda_1 u_1 + \lambda_2 u_2)}(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 \Psi_{u,u_1}(v_1) + \lambda_2 \Psi_{u,u_2}(v_2)$.*

Proof. Let $w = \lambda_1 \Psi_{u,u_1}(v_1) + \lambda_2 \Psi_{u,u_2}(v_2)$. Then by Lemma B.5, $w \in W_{u,(\lambda_1 u_1 + \lambda_2 u_2)}$. On the other hand, by the linearity of $T\exp_N$ we have $T\exp_N(w) = \lambda_1 T\exp_N(\Psi_{u,u_1}(v_1)) + \lambda_2 T\exp_N(\Psi_{u,u_2}(v_2)) = \lambda_1 v_1 + \lambda_2 v_2$. But, by definition, $\Psi_{u,(\lambda_1 u_1 + \lambda_2 u_2)}(\lambda_1 v_1 + \lambda_2 v_2)$ is the unique point in $W_{u,(\lambda_1 u_1 + \lambda_2 u_2)}$ whose image under $T\exp_N$ is $\lambda_1 v_1 + \lambda_2 v_2$, and therefore we must have $w = \Psi_{u,(\lambda_1 u_1 + \lambda_2 u_2)}(\lambda_1 v_1 + \lambda_2 v_2)$, as required. □

Finally, Eliasson [1967] shows that the connection on TN giving rise to $\exp_N$ induces a connection on $T(TN)$, whose corresponding exponential map we denote $\exp_{TN}\colon T(TN) \to TN$. By comparing local expressions, Eliasson [1967] then proves the following crucial lemma:

Lemma B.9. *Let $\exp_{TN}$, $T\exp_N$, and ω be defined as above. Then $\exp_{TN} = T\exp_N \circ \omega$.*

Corollary B.10. *For $u \in T_y N$ sufficiently small, $\exp_{TN}$ is a linear isomorphism between $W_{0,u}$ and $T_{\exp(u)}N$.*

B.3.3. The Structure of the Space of Vector Bundle Maps. We are now in a position to describe the structure on $\mathcal{VB}^{r-1}(TM, TN)$ due to Eliasson [1967]. To do this, let $\mathcal{U}_f$ be a chart for $\mathcal{C}^r(M, N)$ centred at f, as in Section B.1 and denote the bundle above $\mathcal{U}_f$ (i.e., the union of the $\mathcal{VB}^{r-1}_g(TM, TN)$ over $g \in \mathcal{U}_f$) by $\mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN)$. We begin by describing a chart for $\mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN)$ centred on 0_f. By analogy to Section B.1 we want this to take the form $\mathcal{H}_{0_f}\colon \mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN) \to \mathcal{V}_{0_f}$, where $\mathcal{H}_{0_{f^{-1}}}(\zeta) = \exp_{TN} \circ \zeta$ and $\mathcal{V}_{0_f}$ is an open neighbourhood of the origin in an appropriate Banach space of maps from TM to $T(TN)$ satisfying $\tau_{TN} \circ \zeta = 0_f$. However, we also need to impose additional conditions in order to ensure that $\exp_{TN} \circ \zeta \in \mathcal{VB}^{r-1}(TM, TN)$, i.e., that it maps fibres into fibres and is linear on each fibre.

Thus suppose that $G \in \mathcal{VB}^{r-1}_g(TM, TN)$ with $g \in \mathcal{U}_f$ given by $g = \exp_N \circ \eta$, where $\eta \in \mathcal{C}^r(f^* TN)$ and that we have a ζ such that $\exp_{TN} \circ \zeta = G$ and $\tau_{TN} \circ \zeta = 0_f$. Then by Lemma B.9, $T\exp_N(\omega(\zeta(v))) = G(v) \in T_{g(x)}N$. By definition $T\exp_N(T_u(TN)) \subset T_{\exp(u)}N$ for any $u \in TN$, and hence $\exp_N(\tau_{TN}(\omega(\zeta(v)))) = g(x)$ for all $v \in T_xM$. Thus $\tau_{TN}(\omega(\zeta(v))) = \eta(x)$ for all $v \in T_xM$, and since $\tau_{TN} \circ \omega = T\tau_N$, we see that $\zeta(T_xM) \subset (T\tau_N)^{-1}(\eta(x))$, or in other words, ζ maps fibres of TM into fibres of $T\tau_N$. Since $\tau_{TN} \circ \zeta = 0_f$, we therefore have $\zeta(v) \in W_{0,\eta(x)}$, and by Corollary B.9, ζ is given by $\zeta(v) = \omega(\Psi_{\eta(x),0}(G(v)))$. This immediately implies that ζ is linear as a function of v. Conversely we see that if we have a $\zeta\colon TM \to T(TN)$ satisfying $\tau_{TN} \circ \zeta = 0_f$ and mapping fibres of TM linearly into fibres of $T\tau_N$ then $\exp_{TN} \circ \zeta \in \mathcal{VB}^{r-1}(TM, TN)$.

We next need to develop a notation for ζ satisfying such conditions. For later use, it will be convenient to do this in the general case $\tau_{TN} \circ \zeta = F$, rather than just $\tau_{TN} \circ \zeta = 0_f$. With the usual identifications on pull-back bundles, we can denote the space of ζ which map T_xM linearly into $(T\tau_N)^{-1}(\eta(x))$ for a given η by $\mathcal{C}^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$, where $\eta^*(T\tau_N)$ is the pull-back of $T\tau_N\colon T(TN) \to TN$. We shall denote the union of the $\mathcal{C}^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ over all $\eta \in \mathcal{C}^r(f^*TN)$ by $\mathcal{C}^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$. Note that this is not a Banach space since we have no way of adding two functions over different η. However, if we fix some $F \in \mathcal{VB}_f^{r-1}(TM, TN)$ and restrict to those ζ which lie over F (i.e., satisfy $\tau_{TN} \circ \zeta = F$), then we do get a linear structure as we shall see below. We shall denote the space of such maps by $\mathcal{C}_F^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N))) = \{\zeta \in \mathcal{C}^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))\colon \tau_{TN} \circ \zeta = F\}$. If we take local coordinates in $T(TN)$, then maps in this space have the form $\zeta(v) = (f(x), F_x(v), \tilde{\eta}(x), \zeta_x(v))$ where $v \in T_xM$, and F_x, ζ_x are linear maps from T_xM to V such that $F(v) = (f(x), F_x(v))$ and $\eta(x) = (f(x), \tilde{\eta}(x))$. Observe that $\zeta(v) \in W_{F(v),\eta(x)}$.

The vector space structure on $\mathcal{C}_F^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ is that induced from fibres of τ_{TN}. Thus if $\zeta \in \mathcal{C}_F^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$, $\zeta' \in \mathcal{C}_F^{r-1}(\mathcal{L}(TM, (\eta')^*(T\tau_N)))$, and $\lambda, \lambda' \in \mathbb{R}$, then $\zeta(v) \in W_{F(v),\eta(x)}$ and $\zeta'(v) \in W_{F(v),\eta'(x)}$. We can therefore define $\lambda\zeta + \lambda'\zeta'$ using Lemma B.5 by $(\lambda\zeta + \lambda'\zeta')(v) = \lambda\zeta(v) + \lambda'\zeta'(v) \in W_{F(v),(\lambda\eta(x)+\lambda'\eta'(x))}$. We thus see that the definition of $\mathcal{C}_F^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ involves the interplay of two separate linear structures: Whereas the addition of ζ and ζ' uses the structure in $T_{F(v)}(TN)$, the linearity of each ζ is defined in terms of the structure in $(T\tau_N)^{-1}(\eta(x))$. In particular, if $v, v' \in T_xM$, then $\zeta(v), \zeta(v') \in (T\tau_N)^{-1}(\eta(x))$, and hence we form $\lambda\zeta(v) + \lambda'\zeta(v')$ using the addition and scalar multiplication in $(T\tau_N)^{-1}(\eta(x))$.

If we denote the union of $\mathcal{C}_F^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ over $\eta \in \mathcal{V}_f$ by $\mathcal{C}_F^{r-1}(\mathcal{L}(TM, \mathcal{V}_f^*(T\tau_N)))$, we see that the chart for $\mathcal{VB}^{r-1}(TM, TN)$ described above has $\mathcal{V}_{0_f} = \mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, \mathcal{V}_f^*(T\tau_N)))$. Observe that this is a vector bundle over $\mathcal{V}_f$, with the fibre over η being simply $\mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$. Furthermore, we have already shown above that if $\zeta \in \mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ and $\zeta' \in \mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, (\eta')^*(T\tau_N)))$, then $(\lambda\zeta + \lambda'\zeta') \in \mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, (\lambda\eta + \lambda'\eta')^*(T\tau_N)))$. Thus this bundle structure is consistent with the linear structure on $\mathcal{V}_{0_f}$, and can be used to give a trivialization of $\mathcal{V}_{0_f}$. To do this, all that we need is a linear isomorphism between an arbitrary fibre $\mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ and the fibre $\mathcal{C}_{0_f}^{r-1}(\mathcal{L}(TM, 0_f^*(T\tau_N)))$ over $0_f \in \mathcal{C}^r(f^*TN)$ (note that we use the symbol 0_f for both maps in $\mathcal{VB}_f^{r-1}(TM, TN)$ and in $\mathcal{C}^r(f^*TN)$; this slight abuse of notation should cause no difficulties). Such an isomorphism is obtained from the natural isomorphism between $T_u(T_yN)$ and $T_0(T_yN)$ which we shall denote by Ξ_u. In local coordinates we have $\Xi_u(y, \tilde{u}, 0, v_3) = (y, 0, 0, v_3)$. A coordinate-free expression for Ξ_u is thus given by $\Xi_u(\omega) = w - 0_u$, where 0_u is the origin in $T_u(T_y N)$ and the subtraction is carried out using the linear structure in $(T\tau_N)^{-1}(0_y)$ (note that both $T_u(T_y N)$ and $T_0(T_y N)$ lie in $(T\tau_N)^{-1}(0_y)$). This formula is however potentially misleading, since it hides the linearity of Ξ_u with respect to the standard structures in $T_u(T_y N)$ and $T_0(T_y N)$.

An isomorphism between $\mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ and $\mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, 0_f^*(T\tau_N)))$ is then given by $\Xi(\zeta) = \omega \circ \Xi_\eta \circ \omega \circ \zeta$, where Ξ_η is given by $\Xi_{\eta(x)}$ on each fibre $T_{\eta(x)}(T_{f(x)}N)$. This gives the trivialization $\tilde{\Xi}\colon \mathcal{V}_{0_f} \to \mathcal{V}_f \times \mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, 0_f^*(T\tau_N)))$ defined by $\tilde{\Xi}(\zeta) = (\eta, \Xi(\zeta))$. Since $\mathcal{H}_{0_f}$ is fibre preserving and is linear on each fibre of $\mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN)$, we obtain the required trivialization $\tilde{\Xi}$ of $\mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN)$ by conjugating $\tilde{\Xi}$ with $(h_f, \mathcal{H}_{0_f})$. An explicit expression for $\tilde{\Xi}$ is given by $\tilde{\Xi}(G) = (\exp_N, \exp_{TN}) \circ \omega \circ \Xi_\eta \circ \Psi_{\eta,0} \circ G$. It is straightforward, but tedious, to verify that this satisfies the necessary conditions to make $\mathcal{VB}^{r-1}(TM, TN)$ a vector bundle (Eliasson [1967]).

Next, we want to derive an expression for σ' with respect to the above local coordinates. If, as above, $g = \exp_N \circ \eta$, then $\sigma'(g) = T\exp_N \circ T\eta = \exp_{TN} \circ \omega \circ T\eta$. Unfortunately $\tau_{TN} \circ \omega \circ T\eta = T\tau_N \circ T\eta = T(\tau_N \circ \eta) = Tf$, and hence $\omega \circ T\eta \notin \mathcal{V}_{0_f} = \mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ (recall that this is characterized by $\tau_{TN} \circ \zeta = 0_f$). In fact, $\omega \circ T\eta \in \mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \eta^*(T\tau_N)))$. We thus need to construct an appropriate identification between these two spaces. To do this, define $\Theta_{Tf,\eta} \in \mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \eta^*(T\tau_N)))$ by $\Theta_{Tf,\eta}(v) = \omega(\Psi_{\eta(x),Tf(v)}(0_{g(x)}))$, so that $\exp_{TN}(\Theta_{Tf,\eta}(v)) = 0_{g(x)}$. Note that Θ is not quite the same as Eliasson's intertwining operator θ; in fact $\theta(\eta(x))Tf(v) = \vartheta_{Tf,\eta}(v) - \Theta_{Tf,\eta}(v)$, where $\vartheta_{Tf,\eta}(v)$ is the unique point in $W_{Tf(v),\eta(x)} \cap \ker K$ where $K\colon T(TN) \to TN$ is the connection on TN giving rise to $\exp_N$. By Lemma B.9, $(\omega \circ T\eta - \Theta_{Tf,\eta})(v) \in \mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, \eta^*(T\tau_N)))$, and by Lemma B.9 and the linearity of $T\exp_N$ on fibres of τ_{TN} we have $\exp_{TN}(\omega \circ T\eta - \Theta_{Tf,\eta})(v) = \exp_{TN} \circ \omega \circ T\eta(v) - \exp_{TN}(\Theta_{Tf,\eta}(v)) = \sigma'(g) - 0_g = \sigma'(g)$. Hence the local expression for σ' with respect to the chart given by $\mathcal{H}_{0_f}$ is simply $\sigma'(\eta) = \omega \circ T\eta - \Theta_{Tf,\eta}$. If we wish to use the coordinates given by the trivialization of $\mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN)$, we simply compose with $\exp_{TN} \circ \Xi$, giving $\sigma'(\eta) = \exp_{TN} \circ \Xi_\eta \circ (\omega \circ T\eta - \Theta_{Tf,\eta})$. This is essentially the formula given by Eliasson who writes $\pi_0(\omega \circ T\eta) + \theta(\eta(x))Tf(v)$ instead of $\omega \circ T\eta - \Theta_{Tf,\eta}$, where $\pi_0(\omega \circ T\eta) = \omega \circ T\eta - \vartheta_{Tf,\eta}(v)$ is the projection of $\omega \circ T\eta$ onto the vertical subspace $T_{\eta(x)}(T_{g(x)}N)$ obtained from the splitting $T_{\eta(x)}(TN) = T_{\eta(x)}(T_{g(x)}N) \oplus \ker K$. Also Eliasson uses $T\exp_N$ rather than $\exp_{TN}$, thereby eliminating the need for ω (see below).

Unfortunately, this expression is not particularly convenient if we wish to compute $T\sigma'$, since this requires us to calculate the derivative $D_\eta(\omega \circ T\eta - \Theta_{Tf,\eta})$. Whilst the first term is linear in η, there appears to be no way of computing $D_\eta \Theta_{Tf,\eta}$. Fundamentally, this problem arises because $\mathcal{H}_{0_f}$, being centred at 0_f, rather than Tf does not lead to a good model for the tangent space at Tf, i.e., $\mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ does not give a good representation of $T_{Tf}\mathcal{VB}^{r-1}(TM, TN)$. It is thus preferable to seek a chart centred at Tf. We have in fact already carried out all the necessary work to construct such a chart, particularly in Lemma B.7. This implies that given $G \in \mathcal{VB}^{r-1}_g(TM, TN)$ with $g \in \mathcal{U}_f$ there exists a unique $\zeta \in \mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ such that $G = \exp_{TN} \circ \zeta$, and conversely the same argument as for $\mathcal{C}^{r-1}_{0_f}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ shows that if $\zeta \in \mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$, then $\exp_{TN} \circ \zeta \in \mathcal{VB}^{r-1}(TM, TN)$. Hence we have the chart $\mathcal{H}_{Tf}\colon \mathcal{U}_{Tf} \to \mathcal{V}_{Tf}$, where $\mathcal{H}^{-1}_{Tf}(\zeta) = \exp_{TN} \circ \zeta$ and $\mathcal{U}_{Tf} = \mathcal{U}_{0_f} = \mathcal{VB}^{r-1}_{\mathcal{U}_f}(TM, TN)$ are as before, but now $\mathcal{V}_{Tf} = \mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \mathcal{V}^*_f(T\tau_N)))$.

As already observed above, $\omega \circ T\eta \in \mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \mathcal{V}^*_f(T\tau_N)))$, and hence in the coordinates provided by this chart we simply have $\sigma'(\eta) = \omega \circ T\eta$. Since $\omega \circ T\eta$ is linear in η, we have $D_0\sigma'(\eta) = \omega \circ T\eta$, giving the following lemma:

Lemma B.11. *Let $\sigma'\colon \mathcal{C}^{r+1}(M, N) \to \mathcal{C}^r(\mathcal{L}(TM, \mathcal{C}^r(M, N)^*TN))$ be the map $\sigma'(f) = Tf$. Then σ' is $\mathcal{C}^\infty$ and if we identify $T_{Tf}\mathcal{VB}^{r-1}(TM, TN)$ with $\mathcal{C}^{r-1}_{Tf}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ we have*

$$T_f\sigma'(\eta) = \omega \circ T\eta.$$

By combining this with Corollary B.3, we also immediately get the following

Corollary B.12. $T_{f,v}ev_{\sigma'}(\xi, u) = \omega(T_x\eta(v)) + T_v(Tf)(u)$.

Example B.1. It may help to clarify these expressions if we consider the case $N = V$, a Banach space. Then we have natural identifications of TV with $V \times V$, and $T(TV)$ with $V \times V \times V \times V$. The canonical convolution is just $\omega(u_0, u_1, u_2, u_3) = (u_0, u_2, u_1, u_3)$.

The space $\mathcal{C}^r(M, V)$ is itself a Banach space, so that $T\mathcal{C}^r(M, V) = \mathcal{C}^r(M, V) \times \mathcal{C}^r(M, V)$. The tangent space $T_f\mathcal{C}^r(M, V)$ is given by $T_f\mathcal{C}^r(M, V) = \{(f, \tilde{\eta})\colon \tilde{\eta} \in \mathcal{C}^r(M, V)\}$. A vector bundle map $F \in \mathcal{VB}^{r-1}(TM, TV)$ can be written as $F(v) = (f(x), Fx(v))$, where $f \in \mathcal{C}^r(M, V)$ and F_x is a linear map from T_xM to V for each $x \in M$. In particular, we have $\sigma'(f) = (f, T_x f)$. This is linear in f, and hence if we take $\eta = (f, \tilde{\eta}) \in T_f\mathcal{C}^r(M, V)$ we get $T_f\sigma'(\eta) = (f, T_x f, \tilde{\eta}, T_x\tilde{\eta})$. On the other hand, $T\eta = (f, \tilde{\eta}, T_x f, T_x\tilde{\eta})$, and hence $T_f\sigma'(\eta) = \omega \circ T\eta$, as required.

In comparing this calculation to the general case above, the reader may find it helpful to take the standard Riemann structure on V which gives $\exp_V(u_0, u_1) = u_0 + u_1$, $T\exp_V(u_0, u_1, u_2, u_3) = (u_0 + u_1, u_2 + u_3)$ and hence $\exp_{TV}(u_0, u_1, u_2, u_3) = (u_0 + u_2, u_1 + u_3)$. Note that $\exp_V$ is a global isomorphism between $T_{u_0}V$ and V. Also observe that the tangent space $T_F\mathcal{VB}^{r-1}(TM, TV)$ takes the form $T_F\mathcal{VB}^{r-1}(TM, TV) = \{(f, F_x, \tilde{\eta}, \zeta_x)\}$, where ζ_x is linear in v. □

This particular example could be used as the basis of a proof of Lemma B.11 for a much wider class of N, in particular for any N that is an embedded submanifold of a Banach space and has a tubular neighbourhood. This includes all finite-dimensional manifolds, since by the Whitney Embedding Theorem (Hirsch [1976]) any such manifold can be embedded in some Euclidean space. Franks [1979] in turn shows that when N is an embedded submanifold of V and has a tubular neighbourhood, then $\mathcal{C}^r(M, N)$ is an embedded submanifold of $\mathcal{C}^r(M, V)$ (he states the result for finite dimensional N, but his proof merely requires the existence of a tubular neighbourhood). This in fact is how he proves that $\mathcal{C}^r(M, N)$ is a manifold. It seems reasonable to expect that under these conditions $\mathcal{VB}^{r-1}(TM, TN)$ should be an embedded submanifold of $\mathcal{VB}^{r-1}(TM, TV)$. If this is the case, then Example B.1 gives a direct proof of Lemma B.12.

Finally, we point out that in constructing the manifold structure on $\mathcal{VB}^{r-1}(TM, TN)$, we could use $T\exp_N$ to construct the charts, instead of $\exp_{TN}$. This gives an isomorphic structure to that presented here, with the roles of $T\tau_N$ and τ_{TN} interchanged. In particular, the space $\mathcal{C}_F^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T\tau_N)))$ giving the local chart at F is replaced by $\{\zeta \in \mathcal{C}^{r-1}(\mathcal{L}(TM, \mathcal{C}^r(f^*TN)^*(T(TN)))) \colon T\tau_N \circ \zeta = F\}$. It is straightforward to verify that this leads to the formula $T_f\sigma'(\eta) = T\eta$ in Lemma B.12. In the context of Example B.1 this amounts to identifying $T_F\mathcal{VB}^{r-1}(TM, TV)$ with the space $\{(f, \tilde{\eta}, F_x, \zeta_x)\}$ rather than $\{(f, F_x, \tilde{\eta}, \zeta_x)\}$ as above. Clearly, the choice of one or the other of these is a matter of arbitrary convention and has no significant consequences as long as we are consistent in our usage.

Appendix C: Technical Calculations

In this appendix we prove a variety of technical results concerning properties of the mapping $\rho(f, \varphi) = \Phi_{f,\varphi}$ which are used throughout Sections 4–6. Some are simple corollaries of the general results presented in Appendix B, and the remainder use straightforward arguments from elementary differential topology. We make the standing assumption that M is compact.

C.1. Smoothness of ev_ρ

Recall that by Theorem B.2 composition is a smooth mapping. As a simple consequence we get

Lemma C.1. *The map* $\rho_i \colon \mathcal{C}^{2r}(M, M) \to \mathcal{C}^r(M, M)$ *given by* $\rho_i(f) = f^i$ *is* $\mathcal{C}^r$. *If we denote* $x_i = f^i(x)$ *and* $\eta_i = T_f\rho_i(\eta)$, *then* $\eta_0 = 0$ *and*

$$\eta_i(x) = \eta(x_{i-1} + T_{x_1} f(\eta_{i-1}(x)). \tag{C.1.1}$$

Proof. By induction on i. Since $\rho_0(f) = Id$ for all f, ρ_0 is $\mathcal{C}^r$, and $\eta_0 = 0$. Now suppose that ρ_{i-1} for some $i > 1$ is $\mathcal{C}^r$. Since $f^i = f \circ f^{i-1}$ we have $\rho_i(f) = \sigma(\rho_{i-1}(f), f)$ where $\sigma \colon \mathcal{C}^r(M, M) \times \mathcal{C}^{2r}(M, M) \to \mathcal{C}^r(M, M)$ is composition. By Theorem B.2, σ is $\mathcal{C}^r$ and hence by the chain rule ρ_i is $\mathcal{C}^r$. Then η_i is given by $\eta_i = T_f\rho_i(\eta) = T_{\rho_{i-1}(f), f}\sigma(\eta_{i-1}, \eta) = Tf \circ \eta_{i-1} + \eta \circ f^{i-1}$. Evaluating this at x gives the required formula. □

Corollary C.2. *If* ρ_i *is as above, then* ev_{ρ_i} *is* $\mathcal{C}^r$ *and*

$$T_{f,x}ev_{\rho_i}(\eta, 0_x) = \sum_{j=1}^{i} T_{x_j} f^{i-j}(\eta(x_{j-1})) \tag{C.1.2}$$

Proof. Observe that $ev_{\rho_i}(f, x) = \rho_i(f)(x) = ev(x, \rho_i(f))$ where $ev \colon M \times \mathcal{C}^r(M, M) \to M$ is the evaluation function given by $ev(x, f) = f(x)$. Thus by Corollary B.3 and the chain rule ev_{ρ_i} is $\mathcal{C}^r$ and $T_{f,x}ev_{\rho_i}(\eta, v) = \eta_i(x) + T_x f^i(v)$. Thus $T_{f,x}ev_{\rho_i}(\eta, 0_x) = \eta_i(x)$, and (C.1.2) follows by induction. □

In practice (C.1.2) will turn out to be less useful than (C.1.1). We shall never need to evaluate $T_{f,x}ev_{\rho_i}(\eta, v)$ for $v \neq 0$, $i > 1$. Either of the above formulae can be used to compute $T_{f,\varphi,x}ev_\rho$ where $\rho(f, \varphi) = \Phi_{f,\varphi}$. We shall in fact only need to evaluate $T_{f,\varphi,x}ev_\rho$ on vectors of the form $(0_f, \xi, 0_x)$, for $\xi \in T_\varphi\mathcal{C}^{2r}(M, \mathbb{R})$. This is given by a particularly simple expression:

Corollary C.3. *Let $\rho\colon \mathcal{C}^{2r}(M, M) \times \mathcal{C}^{2r}(M, \mathbb{R}) \to \mathcal{C}^r(M, \mathbb{R}^d)$ be given by $\rho(f, \varphi) = \Phi_{f,\varphi}$. Then ev_ρ is $\mathcal{C}^r$ and*

$$T_{f,\varphi,x}ev_r(0_f, x, 0_x) = (\xi(x), \xi(x_1), \dots, \xi(x_{d-1}))^\dagger.$$

Proof. We have $\rho(f, \varphi) = (\varphi \circ \rho_0(f), \dots, \varphi \circ \rho_{d-1}(f))$ and hence ρ is $\mathcal{C}^r$ by Theorem B.2, with $T_{f,\varphi}\rho(0_f, \xi) = (\xi \circ \rho_0(f), \dots, \xi \circ \rho_{d-1}(f))^\dagger + (T_\varphi \circ \eta_0(0_f), \dots, T_\varphi \circ \eta_{d-1}(0_f)) = (\xi \circ \rho_0(f), \dots, \xi \circ \rho_{d-1}(f))^\dagger$. Since $ev_\rho(f, \varphi, x) = \Phi_{f,\varphi}(x) = ev(x, \rho(f, \varphi))$, Corollary B.3 implies that ev_ρ is $\mathcal{C}^r$ and $T_{f,\varphi,x}ev_\rho(0_f, \xi, 0_x) = T_{x,\rho(f,\varphi)}ev(0_x, T_{f,\varphi}\rho(0_f, \xi)) = T_{f,\varphi}\rho(0_f, \xi)(x) = (\xi(x), \xi(x_1), \dots, \xi(x_{d-1}))^\dagger$ as required. □

We next need the analogues of these expressions for skew product systems. Thus define $\tau\colon \mathcal{C}^{2r}(M \times N, M) \to \mathcal{C}^r(M \times N, M^d)$ by

$$\tau(f) = (f^{(0)}, f^{(1)}, f^{(2)}, \dots, f^{(d-1)}),$$

where we recall that $f^{(i+1)} = f \circ (f^{(i)}, g^i)$ and $f^{(0)}(x, y) = x$. Lemma C.1 then gives

Corollary C.4. *Let $\tau_i\colon \mathcal{C}^{2r}(M \times N, M) \to \mathcal{C}^r(M \times N, M)$ be the ith component of t, so that $\tau_i(f) = f^{(i)}$. Then τ_i is $\mathcal{C}^r$, and if $\eta_i = T_f\tau_i(\eta)$, then*

$$\eta_{i+1} = \eta \circ (f^{(i)}, g^i) + Tf \circ (\eta_i, 0),$$

with $\eta_0 = 0$, and hence $\eta_1 = \eta$.

Proof. Observe that $(f^{(i)}, g^i) = (f, g)^i$. Applying Lemma C.1 to (f, g) we immediately see that τ_i is $\mathcal{C}^r$ and $(\eta_{i+1}, 0) = (\eta, 0) \circ (f, g)^i + (Tf, Tg) \circ (\eta_i, 0)$. The first component of this is the required formula. □

Corollary C.5. *The map ev_{τ_i} is $\mathcal{C}^r$ and $T_{f,x,y}ev_{\tau_i}(\eta, 0_x, 0_y)$ is given by*

$$T_{f,x,y}ev_{\tau_i}(\eta, 0_x, 0_y) = \sum_{j=1}^{i} T_{(x_j,y_j)} f^{(i-j)}(\eta(x_{j-1}, y_{j-1}), 0).$$

Proof. This follows by applying Corollary C.2 to (f, g). □

Finally, we get the analogue of Corollary C.3:

Corollary C.6. *Let* ρ: $\mathcal{C}^{2r}(M \times N, M) \times \mathcal{C}^{2r}(N) \times \mathcal{C}^{2r}(M, \mathbb{R}) \to \mathcal{C}^r(M, \mathbb{R}^d)$ *be given by* $\rho(f, g, \varphi) = \Phi_{f,g,\varphi}$. *Then* ev_ρ *is* $\mathcal{C}^r$ *and*

$$T_{f,g,\varphi,x,y}ev_\rho(0_f, 0_g, \xi, 0_x, 0_y) = (\xi(x), \xi(x_1), \dots, \xi(x_{d-1}))^\dagger.$$

C.2. The Tangent Map to the Tangent Operator

Next, we require the analogues of the above formulae for $\rho'(f, \varphi) = T\Phi_{f,\varphi}$. These follow from Lemma B.11. Recall that $\mathcal{VB}^r(TM, TM) = \mathcal{C}^r(\mathcal{L}(TM, \mathcal{C}^r + 1(M, M)^*TM))$ is the space of vector bundle maps between TM and TM, so that if $f \in \mathcal{C}^{r+1}(M, M)$ then $Tf \in \mathcal{VB}^r(TM, TM)$, and that σ': $\mathcal{C}^{r+1}(M, M) \to \mathcal{VB}^r(TM, TM)$ is the map $\sigma'(f) = Tf$.

Lemma C.7. *Let* ρ_i': $\mathcal{C}^{2r+1}(M, M) \to \mathcal{VB}^r(TM, TM)$ *be given by* $\rho_i'(f) = Tf^i$. *Then* ρ_i' *is* $\mathcal{C}^r$. *If we denote* $\eta_i' = T_f\rho_i'(\eta)$ *then* $\eta_1' = \omega \circ T\eta$ *and*

$$\eta_i' = \omega \circ T_\eta \circ T(f^{i-1}) + \omega \circ T(Tf) \circ \omega \circ \eta_{i-1}'.$$

Proof. We have $\rho_i' = \sigma' \circ \rho_i$ with ρ_i as above. By Lemma B.11, σ' is $\mathcal{C}^\infty$, whilst by Lemma C.1, ρ_i is $\mathcal{C}^r$ and hence so is ρ_i'. Then $\eta_i' = T_f\rho_i'(\eta) = T\sigma' \circ T_f\rho_i(\eta) = \omega \circ T\eta_i$. Since $\eta_1 = \eta$ by Lemma C.1, this immediately gives $\eta_1' = \omega \circ T_\eta$, as required. Also by Lemma C.1 we have $\omega \circ T\eta_i = \omega \circ T(\eta \circ f^{i-1} + Tf \circ \eta_{i-1}) = \omega \circ T_\eta \circ T(f^{i-1}) + \omega \circ T(Tf) \circ \omega \circ \omega \circ T\eta_{i-1} = \omega \circ T\eta \circ T(f^{i-1}) + \omega \circ T(Tf) \circ \omega \circ \eta_{i-1}'$, as required. □

Corollary C.8. *Let* ρ': $\mathcal{D}^{2r+1}(M, M) \times \mathcal{C}^{2r+1}(M, \mathbb{R}) \to \mathcal{C}^r(\tilde{T}M, T\mathbb{R}^d)$ *be given by* $\rho'(f, \varphi) = T\Phi_{f,\varphi}$. *Then* $ev_{\rho'}$ *is* $\mathcal{C}^r$ *and*

$$T_{f,\varphi,v}ev_{\rho'}(0_f, \xi, 0_v) = (\omega(T_x\xi(v)), \omega(T_{x_1}\xi(v_1)), \dots, \omega(T_{x_{i-1}}\xi(v_{i-1})))^\dagger,$$

where $x_i = f^i(x)$ *and* $v_i = T_xf^i(v)$.

Proof. We have $\rho'(f, \varphi) = (\sigma'(\varphi) \circ \rho_0'(f), \dots, \sigma'(\varphi) \circ \rho_{d-1}'(f))$. By Lemma B.11, σ' is $\mathcal{C}^\infty$, and hence by Theorem B.2 and Lemma C.7, ρ' is $\mathcal{C}^r$, with $T_{f,\varphi}\rho'(0_f, \xi) = (\omega \circ T\xi \circ \rho_0'(f), \dots, \omega \circ T\xi \circ \rho_{d-1}'(f))^\dagger + (T(T_\varphi) \circ \eta_0'(0_f), \dots, T(T_\varphi) \circ \eta_{d-1}'(0_f)) = (\omega \circ T\xi \circ \rho_0'(f), \dots, \omega \circ T\xi \circ \rho_{d-1}'(f))^\dagger$. Since $ev_{\rho'}(f, \varphi, v) = T\Phi_{f,\varphi}(v) = ev(v, \rho'(f, \varphi))$, and $\tilde{T}M$ is compact, Corollary B.3 implies that $ev_{\rho'}$ is $\mathcal{C}^r$ and $T_{f,\varphi,v}ev_{\rho'}(0_f, \xi, 0_v) = T_{v,\rho'(f,\varphi)}ev(0_v, T_{f,\varphi}\rho'(0_f, \xi)) = T_{f,\varphi}\rho'(0_f, \xi)(v) = ((\omega \circ T\xi \circ \rho_0'(f), \dots, \omega \circ T\xi \circ \rho_{d-1}'(f))(v))^\dagger = (\omega(T_x\xi(v)), \omega(T_{x_1}\xi(v_1)), \dots, \omega(T_{x_{i-1}}\xi(v_{i-1})))^\dagger$, as required. □

Finally, as in Section C.1, we derive the skew product analogue of Lemma C.7. Define the map τ': $\mathcal{C}^{2r+1}(M \times N, M) \to \mathcal{VB}^r(T(M \times N), TM^d)$ by

$$\tau'(f) = (Tf^{(0)}, Tf^{(1)}, Tf^{(2)}, \dots, Tf^{(d-1)}),$$

and denote its ith component by τ_i'. Then

Lemma C.9. *The map τ' is $\mathcal{C}^r$. Let $\eta'_i = T_f \tau'_i(\eta)$, then $\eta'_0 = 0$ and*

$$\eta'_{i+1} = \omega \circ T_\eta \circ T(f^{(i)}, g^i) + \omega \circ T(Tf) \circ (\omega \circ \eta'_i, 0),$$

where as in Corollary C.4, $\eta_i = T_f \tau_i(\eta)$ with $\tau_i(f) = f^{(i)}$. Furthermore, $ev_{\tau'_i}$ is $\mathcal{C}^r$ and $T_{f,v}(ev_{\tau'_i})(\eta, 0_v) = \eta'_i(v)$.

Proof. Applying Lemma C.7 to $(f^{(i)}, g^i) = (f, g)^i$ we get $(\eta'_i, 0) = \omega \circ T(\eta, 0) \circ T(f^{(i)}, g^i) + \omega \circ T(T(f, g)) \circ \omega \circ (\eta'_{i-1}, 0)$. Taking the first component (and using the same symbol for the canonical convolution on $T(T(M \times N))$ and $T(TM)$, gives the required expression. Then, by definition $ev_{\tau'_i}(f, v) = ev(v, \tau'_i(f))$ where $ev\colon \tilde{T}(M \times N) \times \mathcal{C}^r(\tilde{T}(M \times N), TM) \to TM$ is the evaluation function given by $ev(v, F) = F(v)$. Thus by Corollary B.3 and the chain rule, $ev_{\tau'_i}$ is $\mathcal{C}^k$ and $T_{f,v}(ev_{\tau'_i})(\eta, 0_v) = \eta'_i(v)$ as required. □

C.3. Submersivity

In order to prove the transversality of various evaluation maps we shall often need to show the existence of suitable functions in $T_f\mathcal{D}^r(M)$ and $T_\varphi\mathcal{C}^r(M, \mathbb{R})$. In particular, we shall require the submersivity of the maps $f \mapsto (f(x_0), \ldots, f(x_{j-1}))$ and $f \mapsto (T_{x_0} f, \ldots, T_{x_{j-1}} f)$ when $x_0, \ldots, x_{j-1}$ are disjoint points. These are all based on the following simple lemma.

Lemma C.10. *Let U_0 and U_1 be open neighbourhoods of some $x \in M$, with $\bar{U}_1 \subset U_0$, where $\bar{U}_1$ denotes the closure of U_1. Then there exists a $\mathcal{C}^\infty$ function $\psi\colon M \to [0, 1]$ such that $\psi \equiv 1$ on U_1 and the support of ψ is contained in U_0.*

For a proof see for instance Hirsch [1976]. Straightforward applications of this gives

Lemma C.11. *Let $f \in \mathcal{C}^r(M, N)$. Given any $x \in M$, $v \in T_{f(x)}N$, and $U \subset M$ an open neighbourhood of x, there exists a $\mathcal{C}^r$ function $\eta \in \mathcal{C}^r(f^*TN) = T_f\mathcal{C}^r(M, N)$ such that $\eta(x) = v$ and the support of η is contained in U.*

Proof. Let $(\tau_N^{-1}(V), \alpha)$ be a local trivialization of TN for some $V \subset N$ with $f(x) \in V$. Suppose that $\alpha(v) = (f(x), b)$ for some $b \in \mathbb{R}^n$. Define $\eta_0\colon M \to TN$ by

$$\eta_0(y) = \begin{cases} \alpha^{-1}(f(y), b), & y \in f^{-1}(V), \\ 0_{f(y)}, & \text{otherwise.} \end{cases}$$

Using the lemma, choose $\psi\colon M \to [0, 1]$ such that $\psi(x) = 1$ and the support of ψ is contained in $f^{-1}(V) \cap U$. Let $\eta\colon M \to TN$ be given by $\eta(y) = \psi(y)\eta_0(y)$. Since ψ is identically zero in an open neighbourhood of any point in the boundary of $f^{-1}(V)$, η is $\mathcal{C}^r$. Furthermore the support of η is contained in U, and $\eta(x) = \psi(x)\eta_0(x) = \alpha^{-1}(x, b) = v$, as required. □

Corollary C.12. *Given any finite set of distinct points* $\{x_0, \ldots, x_{j-1}\} \subset M$ *and any set of* $v_i \in T_{f(x_i)}N$, $i = 0, \ldots, j-1$, *we can find a* $\eta \in \mathcal{C}^r(f^*TN)$ *such that* $\eta(x_i) = v_i$ *for all* $i = 0, \ldots, j-1$.

Proof. For each $i = 0, \ldots, j-1$, choose a neighbourhood U_i of x_i such that $x_j \notin U_i$ if $i \neq j$. Construct $\eta_i \in \mathcal{C}^r(f^*TN)$ such that $\eta_i(x_i) = v_i$ and the support of η_i is contained in U_i. Then $\eta = \eta_0 + \cdots + \eta_{j-1}$ has the required properties. □

Corollary C.13. *If* $x_0, \ldots, x_{j-1}$ *are distinct, then the map* $\rho\colon \mathcal{C}^r(M, N) \to N^j$ *given by* $\rho(f) = (f(x_0), \ldots, f(x_{j-1}))$ *is a submersion.*

Proof. Since $T_f\rho(\eta) = (\eta(x_0), \ldots, \eta(x_{j-1}))$, this is just a restatement of the Corollary C.12. □

The above two corollaries show that we can choose η independently on $x_0, \ldots, x_{j-1}$. We shall need essentially the same result for $T\eta$. It shall be useful to give two forms of this result, both based on Lemma C.14 below. First, some notation.

Given any $x \in M$, $y \in N$, let (U, β) be a chart centred at x and $((\tau_N)^{-1}(V), \alpha)$ a local trivialization for TN, with $y \in V$. Thus $\beta\colon U \to \mathbb{R}^n$ with $\beta(x) = 0$, and $\alpha\colon (\tau_N)^{-1}(V) \to V \times \mathbb{R}^n$ such that $\pi_1 \circ \alpha = \tau_N$, and α is a linear isomorphism on each fibre, where $\pi_1 : V \times \mathbb{R}^n \to V$ is the projection onto the first factor. Denote the restriction of α to T_yN by α_y; note that this gives an isomorphism between T_yN and $\mathbb{R}^n$. Finally let $\alpha'\colon T\mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ be a (global) trivialization for $T\mathbb{R}^n$.

Let $\mathcal{L}(T_xM, T_yN)$ be the space linear maps $T_xM \to T_yN$. Note that T_xM consists of equivalence classes $[U, \beta, v']$ with $v' \in \mathbb{R}^m$. This allows us to identify T_xM and $\mathbb{R}^m$, via $\beta'(v') = [U, \beta, v']$. Hence given any $B \in \mathcal{L}(T_xM, T_yN)$ and any $b \in \mathbb{R}^n$, we get a map $\zeta_{b,B} = b + \alpha_y \circ B \circ \beta' \circ \beta\colon U \to \mathbb{R}^n$. Since $\beta(x) = 0$, we have $\zeta_{b,B}(x) = b$.

If $f \in \mathcal{C}^r(M, N)$ such that $f(x) = y$, define $\eta_{b,B} \in T_f\mathcal{C}^r(M, N)$ by

$$\eta_{b,B}(z) = \begin{cases} \psi(z)\alpha^{-1}(f(z), \zeta_{b,B}(z)), & z \in U \cap f^{-1}(V), \\ \psi(z)\alpha^{-1}(f(z), 0), & \text{otherwise}, \end{cases} \tag{C.3.1}$$

where ψ is identically 1 in a neighbourhood of x and has support contained in $f^{-1}(V) \cap U$. Then $\eta_{b,B}$ has support contained in U and $\eta_{b,B}(x) = \alpha^{-1}(x, b)$.

Lemma C.14. *With* $\eta_{b,B}$ *defined as above, we have for any* $v \in T_xM$

$$T_x\eta_{b,B}(v) = T\alpha^{-1}(T_xf(v), (\alpha')^{-1}(b, \alpha_y \circ B(v))).$$

Proof. By definition if $g\colon U \to \mathbb{R}^n$, then $T_xg(v) = (\alpha')^{-1}(g(x), D_x(g \circ \beta^{-1}).v')$, where $v = [U, \beta, v'] \in T_xM$. Thus $\alpha' \circ T_x(\zeta_{b,B})(v) = (\zeta_{b,B}(x), D_x(b + \alpha_y \circ B \circ b').v')$. Since $\alpha_y \circ B \circ b'$ is linear, we have $D_x(\alpha_y \circ B \circ \beta').v' = \alpha_y \circ B \circ \beta'(v') = \alpha_y \circ B(v)$. Thus $\alpha'(T_x(\zeta_{b,B})(v)) = (b, \alpha_y \circ B(v))$, as required. □

Corollary C.15. *Given any* $v \in T_xM$ *with* $v \neq 0$ *and any* $w \in T_u(TN)$ *where* $u = T_xf(v)$, *and an open neighbourhood and* $U \subset M$ *of* x, *there exists* $\eta \in T_f\mathcal{C}^r(M, N)$ *such that* $\omega(T_x\eta(v)) = w$, *and the support of* η *is contained in* U.

Proof. Recall that if ω is the canonical involution on $T(TN)$, then $T\tau_N \circ \omega = \tau_{TN}$ and $\omega \circ \omega = Id$. Thus, since $\pi_1 \circ \alpha = \tau_N$, we have $T\pi_1 \circ T\alpha = T\tau_N$ and hence $T\pi_1 \circ T\alpha \circ \omega = \tau_{TN}$. By definition $\tau_{TN}(w) = u$ and so $T\alpha(\omega(w)) = (u, (\alpha')^{-1}(w_2, w_3))$ for some $w_2, w_3 \in \mathbb{R}^n$. Thus $w = \omega(T\alpha^{-1}(u, (\alpha')^{-1}(w_2, w_3)))$. Choose $b = w_2$ and B to be any map in $\mathcal{L}(T_x M, T_y N)$ such that $\alpha_y(B(v)) = w_3$; this is possible since $v \neq 0$ and α_y is an isomorphism. If necessary, shrink U so that (U, β) is a chart for some $\beta\colon U \to \mathbb{R}^m$. Then if we set $\eta = \eta_{b,B}$ we have as required,

$$\begin{aligned} \omega(T_x\eta(v)) &= \omega(T\alpha^{-1}(T_x f(v), (\alpha')^{-1}(b, \alpha_y \circ B(v)))) \\ &= \omega(T\alpha^{-1}(u, (\alpha')^{-1}(w_2, w_3))) \\ &= w. \end{aligned}$$

□

Corollary C.16. *Given any finite set of distinct points* $\{x_0, \dots, x_{j-1}\} \subset M$*, any set of* $v_i \in T_{x_i} M$*,* $v_i \neq 0$*, and any collection of* $w_i \in T_{u_i}(TN)$ *where* $u_i = T_{x_i} f(v_i)$*, we can find an* $\eta \in \mathcal{C}^r(f^*TN)$ *such that* $\omega(T_{x_i}\eta(u_i)) = w_i$ *for all* $i = 0, \dots, j-1$*.*

Proof. Exactly analogous to the proof of Corollary C.12: For each $i = 0, \dots, j-1$ construct η_i such that $\omega(T_{x_i}\eta_i(u_i)) = w_i$ and x_j is not in the support of η_i if $i \neq j$, then $\eta = \eta_0 + \cdots + \eta_{j-1}$ has the required properties. □

The above corollary is all we shall need in proving the immersivity of $\Phi_{f,\varphi}$ at general points of M. However, at periodic points, it will be more convenient to restrict to perturbations that preserve the position of the periodic point. So, given any $x \in M$, $y \in N$, define $\mathcal{C}^r(M, N; x; y) = \{f \in \mathcal{C}^r(M, N)\colon\ f(x) = y\}$. If $\rho\colon \mathcal{C}^r(M, N) \to N$ is given by $\rho(f) = f(x)$, then $\mathcal{C}^r(M, N; x; y) = \rho^{-1}(\{y\})$ and since by Corollary C.13, ρ is a submersion, $\mathcal{C}^r(M, N; x; y)$ is a submanifold of $\mathcal{C}^r(M, N)$. Furthermore, $T_f\mathcal{C}^r(M, N; x; y) = T\rho^{-1}(0) = \{\eta \in T_f\mathcal{C}^r(M, N)\colon\ \eta(x) = 0\}$.

For any $A \in \mathcal{L}(T_x M, T_y N)$ we may identify $T_A\mathcal{L}(T_x M, T_y N)$ with those maps $\tilde{A}$ in $\mathcal{L}(T_x M, T(T_y N))$ such that $\tau_{TN} \circ \tilde{A} = A$. This gives $T_A\mathcal{L}(T_x M, T_y N)$ a structure compatible with that defined on $\mathcal{C}^r(\mathcal{L}(TM, \mathcal{C}^r(M, N)^*TN))$ in Appendix B.3.

Furthermore, if $\eta \in T_f\mathcal{C}^r(M, N)$, we automatically have $\tau_{TN}(\omega \circ T_x\eta) = T_x f$ since $\tau_N \circ \eta = f$ and $\tau_{TN} \circ \omega = T\tau_N$. Also if $\eta(x) = 0_y$ then $\tau_{TN}(T_x\eta(v)) = 0_y$ for all $v \in T_x M$, and hence $T\tau_N(\omega(T_x\eta(v))) = 0_y$. But $T_y N = (\tau_N)^{-1}(y)$, and hence $T(T_y N) = (T\tau_N)^{-1}(0_y)$. Thus $\omega(T_x\eta(v)) \in T(T_y N)$ for all $v \in T_x M$. We have thus shown that if $\eta \in T_f\mathcal{C}^r(M, N; x; y)$ then $\omega \circ T_x\eta \in T_{T_x f}\mathcal{L}(T_x M, T_y N)$.

Recall that α_y is an isomorphism between $T_y N$ and $\mathbb{R}^n$, and $\alpha'\colon\ T\mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ is a trivialization for $T\mathbb{R}^n$. Then $(\alpha_y^{-1} \times \alpha_y^{-1}) \circ \alpha' \circ T\alpha_y\colon\ T(T_y N) \to T_y N \times T_y N$ gives an induced trivialization for $T(T_y N)$. Define $T_i\alpha = \alpha_y^{-1} \circ \pi_i \circ \alpha' \circ T\alpha_y$ for $i = 1, 2$, where $\pi_1, \pi_2\colon\ \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ are the two projections. Note that $T_1\alpha$ is just the bundle projection $\tau_{TN} : T(T_y N) \to T_y N$. Thus if $\tilde{A} \in T_A\mathcal{L}(T_x M, T_y N)$, then $T_1\alpha \circ \tilde{A} = A$ and $\tilde{A} \mapsto T_2\alpha \circ \tilde{A}$ is a (linear) isomorphism between $T_A\mathcal{L}(T_x M, T_y N)$ and $\mathcal{L}(T_x M, T_y N)$. We then get a further corollary to Lemma C.14:

Corollary C.17. *Given any $x \in M, y \in N$, let ρ: $f \in \mathcal{C}^r(M, N; x; y) \to \mathcal{L}(T_x M, T_y N)$ be defined by $\rho(f) = T_x f$. Then given any $\tilde{A} \in T_{T_x f}\mathcal{L}(T_x M, T_y N)$ and an open neighbourhood U of x, there exists a $\eta \in T_f\mathcal{C}^r(M, N; x; y)$ such that $T_f\rho(\eta) = \tilde{A}$ and the support of η is contained in U.*

Proof. By the discussion above, if $\eta \in T_f\mathcal{C}^r(M, N)$ and $\eta(x) = 0_y$, then automatically $\omega \circ T_x\eta \in T_{T_x f}\mathcal{L}(T_x M, T_y N)$. Since $\tilde{A} \mapsto T_2\alpha \circ \tilde{A}$ is an isomorphism it thus suffices to show that we can construct such a η so that $T_2\alpha \circ T_f\rho(\eta) = T_2\alpha \circ \tilde{A}$ (and whose support is contained in U). As in Corollary C.15, if necessary shrink U so that (U, β) is a chart for some β: $U \to \mathbb{R}^m$ and define $\eta = \eta_{b,B}$ with $b = 0$ and $B = T_2\alpha \circ \tilde{A} \in \mathcal{L}(T_x M, T_y N)$.

By Lemma B.11, and Lemma C.14 we have $T_f\rho(\eta)(v) = \omega(T_x\eta(v)) = \omega(T\alpha^{-1}(T_x f(v), (\alpha')^{-1}(b, \alpha_y \circ B(v))))$ for any $v \in T_x M$. Since α is a trivialization in the neighbourhood of y, we have $T_x f(v) = \alpha^{-1}(y, u_1)$ for some $u_1 \in \mathbb{R}^n$, and $0_y = \alpha^{-1}(y, 0)$. Direct calculation shows that for any $u_1, u_2, u_3 \in \mathbb{R}^n$ we have $T\alpha \circ \omega \circ T\alpha^{-1}(\alpha^{-1}(y, u_1), (\alpha')^{-1}(u_2, u_3)) = (\alpha^{-1}(y, u_2), ((\alpha')^{-1}(u_1, u_3)))$ and $T\alpha_y \circ T\alpha^{-1}(0_y, (\alpha')^{-1}(u_1, u_3)) = (\alpha')^{-1}(u_1, u_3)$. Thus,

$$
\begin{aligned}
T_2\alpha \circ T_f\rho(\eta)(v) &= \alpha_y^{-1} \circ \pi_2 \circ \alpha' \circ T\alpha_y \circ \omega \circ T_x\eta(v) \\
&= \alpha_y^{-1} \circ \pi_2 \circ \alpha' \circ T\alpha_y \circ \omega(T\alpha^{-1}(T_x f(v), (\alpha')^{-1}(0, \alpha_y \circ B(v)))) \\
&= \alpha_y^{-1} \circ \pi_2 \circ \alpha' \circ T\alpha_y \circ T\alpha^{-1}(\alpha^{-1}(y, 0), (\alpha')^{-1}(u_1, \alpha_y \circ B(v))) \\
&= \alpha_y^{-1} \circ \pi_2 \circ \alpha' \circ (\alpha')^{-1}(u_1, \alpha_y \circ B(v)) \\
&= \alpha_y^{-1} \circ \pi_2(u_1, \alpha_y \circ B(v)) \\
&= B(v) \\
&= T_2\alpha \circ \tilde{A}(v),
\end{aligned}
$$

and hence $T_f\rho(\eta) = \tilde{A}$, as required. □

Finally, we get the usual extension of this result to any finite set of distinct points. Thus given $\{x_0, \ldots, x_{j-1}\} \subset M$ as above, and any set of points $\{y_0, \ldots, y_{j-1}\} \subset N$, we generalize $\mathcal{C}^r(M, N; x; y)$ to

$$
\mathcal{C}^r(M, N; x_0, \ldots, x_{j-1}; y_0, \ldots, y_{j-1}) = \{f \in \mathcal{C}^r(M, N)\colon\ f(x_i) = y_i \text{ for all } i = 0, \ldots, j-1\}.
$$

Then, as before, $\mathcal{C}^r(M, N; x_0, \ldots, x_{j-1}; y_0, \ldots, y_{j-1})$ is a submanifold of $\mathcal{C}^r(M, N)$ and its tangent space at f is given by those $\eta \in T_f\mathcal{C}^r(M, N)$ such that $\eta(x_i) = 0$ for all $i = 0, \ldots, j-1$. An analogous argument to Corollaries C.8 and C.12 gives the following.

Corollary C.18. *Suppose that $\{x_0, \ldots, x_{j-1}\} \subset M$ are distinct. Then the map ρ_j: $\mathcal{C}^r(M, N; x_0, \ldots, x_{j-1}; y_0, \ldots, y_{j-1}) \to \mathcal{L}(T_{x_0}M, T_{y_0}N) \times \cdots \times \mathcal{L}(T_{x_{j-1}}M, T_{y_{j-1}}N)$ given by $\rho_j(f) = (T_{x_0}f, \ldots, T_{x_{j-1}}f)$ is a submersion.*

Acknowledgements

The results presented here were motivated by a question asked by David Broomhead at the IUTAM conference held at UCL in July 1993. Subsequently, conversations with Ludwig Arnold, Colin Sparrow, Jerry Huke, and Mike Davies were absolutely crucial to its development, and I am extremely grateful for all their help. Some of the preliminary work on this paper was carried out at IMPA, Brazil, during the International Conference on Dynamical Systems held there in August 1993, and I am extremely grateful to IMPA for their hospitality and financial support, and to the Royal Society for providing travel funds to attend this conference. I would also like to thank the Basic Research Institute in Mathematical Sciences at Hewlett-Packard in Bristol for its hospitality in the summer of 1994. Finally, I would like to gratefully acknowledge financial support from the UK Engineering and Physical Sciences Research Council (GR/J97243) and from the Nuffield Foundation (SCI/180/92/365/G).

References

[1] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, and L. S. Tsimring, 1993, The Analysis of Observed Chaotic Data in Physical Systems, *Rev. Mod. Phys.*, **65**, 1331–1392.

[2] R. Abraham, 1963, Transversality in Manifolds of Mappings, *Bull. Am. Math. Soc.*, **69**, 470–474.

[3] R. Abraham and J. Robbin, 1967, *Transversal Mappings and Flows*, W. D. Benjamin, New York.

[4] D. Aeyels, 1981, Generic Observability of Differentiable Systems, *SIAM J. Control Optim.*, **19**, 595–603.

[5] R. Badii, G. Broggi, B. Derighetti, M. Ravani, S. Cilberto, A. Politi, and M. A. Rubio, 1988, Dimension Increase in Filtered Chaotic Signals, *Phys. Rev. Lett.*, **60**, 979–982.

[6] M. F. Barnsley, 1988, *Fractals Everywhere*, Academic Press, Boston.

[7] S. A. Billings, M. J. Korenberg, and S. Chen, 1988, Identification of Nonlinear Output-Affine Systems Using an Orthogonal Least Squares Algorithm, *Int. J. Syst. Sci.*, **19**, 1559–1568.

[8] D. S. Broomhead, J. P. Huke, and M. R. Muldoon, 1992, Linear Filters and Non-linear Systems, *J. Roy. Stat. Soc. B*, **54**, 373–382.

[9] M. Casdagli, 1992, A Dynamical Systems Approach to Modelling Input-Output Systems, in *Nonlinear Modelling and Forecasting*, ed. M. Casdagli and S. Eubank, Addison-Wesley, Reading, MA.

[10] K. M. Cuomo and A. V. Oppenheim, 1993, Circuit Implementation of Synchronized Chaos with Applications to Communication, *Phys. Rev. Lett.*, **71**, 65–68 (reprinted in Ott et al., 1994).

[11] W. Ditto, S. N. Rauseo, and M. L. Spano, 1990, Experimental Control of Chaos, *Phys. Rev. Lett.*, **65**, 3211–3214 (reprinted in Ott et al., 1994).

[12] J.-P. Eckmann and D. Ruelle, 1985, Ergodic Theory of Chaos and Strange Attractors, *Rev. Mod. Phys.*, **57**, 617–656.

[13] J. Eells, 1958, On the Geometry of Function Spaces, *Symposium Internacional de Topologiá Algebraica*, 303–308, Universidad Nacional Autónoma de México and UNESCO, Mexico City.

[14] J. Eells, 1966, A Setting for Global Analysis, *Bull. Am. Math. Soc.*, **72**, 751–807.

[15] H. I. Eliasson, 1967, Geometry of Manifolds of Maps, *J. Diff. Geom.*, **1**, 169–194.

[16] F. Flepp, R. Holzner, E. Brun, M. Finardi, and R. Badii, 1991, Model Identification by Periodic-Orbit Analysis for NMR-Laser Chaos, *Phys. Rev. Lett.*, **67**, 2244–2247 (reprinted in Ott et al., 1994).

[17] M. J. Foster, 1975, Calculus on Vector Bundles, *J. London Math. Soc.*, **11**, 65–73.
[18] J. Franks, 1979, Manifolds of $\mathcal{C}^r$ Mappings and Applications to Dynamical Systems, *Adv. Math. Suppl. Stud.*, **4**, 271–290.
[19] Z. Gills, C. Iwata, R. Roy, I. Schwartz, and I. Triandaf, 1992, Tracking Unstable Steady States: Extending the Stability Range of a Multi-Mode Laser, *Phys. Rev. Lett.*, **69**, 3169–3172 (reprinted in Ott et al., 1994).
[20] P. Grassberger, T. Schreiber, and C. Schaffrath, 1992, Nonlinear Time Sequence Analysis, *Int. J. Bifurcation Chaos*, **1**, 521–547.
[21] B. T. Grenfell, 1992, Chaos and Chance in Measles Dynamics, *J. Roy. Stat. Soc. B*, **54**, 383–398.
[22] S. Hayes, C. Grebogi, and E. Ott, 1993, Communicating with Chaos, *Phys. Rev. Lett.*, **70**, 3031–3034 (reprinted in Ott et al., 1994).
[23] M. Hirsch, 1976, *Differential Topology*, Springer-Verlag, New York.
[24] J. P. Huke, 1993, Embedding Nonlinear Dynamical Systems, A Guide to Takens Theorem, Internal Report, DRA Malvern.
[25] E. R. Hunt, 1991, Stabilizing High-Period Orbits in a Chaotic System: The Diode Resonator, *Phys. Rev. Lett.*, **67**, 1953–1955 (reprinted in Ott et al., 1994).
[26] M. C. Irwin, 1980, *Smooth Dynamical Systems*, Academic Press, London.
[27] Y. Kifer, 1988, *Random Perturbations of Dynamical Systems*, Birkhäuser, Boston.
[28] S. Lang, 1972, *Differential Manifolds*, Addison-Wesley, Reading, MA.
[29] F. C. Moon and P. J. Holmes, 1979, A Magnetoelastic Strange Attractor, *J. Sound Vib.*, **65**, 285–296.
[30] L. Noakes, 1991, The Takens Embedding Theorem, *Int. J. Bifurcation Chaos*, **1**, 867–872.
[31] M. F. Norman, 1968, Some Convergence Theorems for Stochastic Learning Models with Distance Diminishing Operators, *J. Math. Psychology*, **5**, 61–101.
[32] E. Ott, T. Sauer, and J. A. Yorke, 1994, *Coping with Chaos*, Wiley, New York.
[33] R. S. Palais, 1968, *Foundations of Global Nonlinear Analysis*, W. A. Benjamin, New York.
[34] J. Palis and W. de Melo, 1982, *Geometric Theory of Dynamical Systems*, Springer-Verlag, New York.
[35] F. Papoff, A. Fioretti, E. Arimondo, G. B. Mindlin, H. Solari, and R. Gilmore, 1992, Structure of Chaos in the Laser with Saturable Absorber, *Phys. Rev. Lett.*, **68**, 1128–1131 (reprinted in Ott et al., 1994).
[36] L. M. Pecora and T. L. Carroll, 1990, Synchronisation in Chaotic System, *Phys. Rev. Lett.*, **64**, 821–824 (reprinted in Ott et al., 1994).
[37] L. M. Pecora and T. L. Carroll, 1991, Driving Systems with Chaotic Signals, *Phys. Rev. A*, **44**, 2374–2383.
[38] F. Quinn, 1970, Transversal Approximation on Banach Manifolds, *Global Analysis*, ed. S.-S. Chern and S. Smale, *AMS Proc. Symp. Pure Math.*, **XV**, 213–222.
[39] D. Ruelle, 1989, *Elements of Differentiable Dynamics and Bifurcation Theory*, Academic Press, Boston.
[40] F. J. Romeiras, A. Bondeson, E. Ott, T. M. Antonsen, and C. Grebogi, 1987, Quasi-periodically Forced Dynamical Systems with Strange Nonchaotic Attractors, *Physica D*, **26**, 277–294.
[41] T. Sauer, J. A. Yorke, and M. Casdagli, 1991, Embedology, *J. Stat. Phys.*, **65**, 579–616.
[42] T. Shinbrot, W. L. Ditto, C. Crebogi, E. Ott, M. L. Spano, and J. A. Yorke, 1992, Using the Sensitive Dependence of Chaos (the "Butterfly Effect") to Direct Trajectories in an Experimental Chaotic System, *Phys. Rev. Lett.*, **68**, 2863–2868 (reprinted in Ott et al., 1994).
[43] S. Smale, 1963, Stable Manifolds for Differential Equations and Diffeomorphisms, *Ann. Sc. Norm. Sup. Pisa*, **18**, 97–116.
[44] S. Smale, 1965, An Infinite Dimensional Version of Sard's Theorem, *Am. J. Math.*, **87**, 861–866.
[45] J. C. Sommerer, W. L. Ditto, C. Crebogi, E. Ott, and M. L. Spano, 1991, Experimental Confirmation of the Theory for Critical Exponents of Crises, *Phys. Lett. A*, **153**, 105–109 (reprinted in Ott et al., 1994).
[46] J. Stark, 1993, Recursive Prediction of Chaotic Time Series, *J. Nonlin. Sci.*, **3**, 197–223.

[47] J. Stark and M. E. Davies, 1994, Recursive Filters Driven by Chaotic Signals, *IEE Colloquium on Exploiting Chaos in Signal Processing, IEE Digest* **143**, 5/1–5/16.
[48] F. Takens, 1980, Detecting Strange Attractors in Turbulence, in *Dynamical Systems and Turbulence, Warwick, 1980*, ed. D. A. Rand and L.-S. Young, Lecture Notes in Mathematics, **898**, Springer-Verlag, Berlin.
[49] R. Thom, 1954, Quelques Propriétés Globales des Variétes Différentiables, *Commun. Math. Helv.*, **28**, 17–86.
[50] R. Thom, 1956, Une Lemma sur les Applications Différentiables, *Bol. Soc. Math. Mex.*, **1**, 59–71.
[51] N. B. Tufillaro, P. Wyckoff, R. Brown, T. Schreiber, and T. Molteno, 1995, Topological Time Series Analysis of a String Experiment and its Synchronized Model, *Phys. Rev. E*, **51**, 164–174.